2011-06

Prioritizing unaided human search in military simulations

Starling, James Kendall.

Monterey, California. Naval Postgraduate School

NAVAL
POSTGRADUATE
SCHOOL

MONTEREY, CALIFORNIA

THESIS

PRIORITIZING UNAIDED HUMAN SEARCH IN MILITARY SIMULATIONS

by

James Starling

June 2011

Thesis Advisor: Carlos F. Borges
Second Reader: Paul F. Evangelista

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	June 2011	Master's Thesis – June 09 - June 11

4. TITLE AND SUBTITLE: Prioritizing Unaided Human Search in Military Simulations

5. FUNDING NUMBERS

6. AUTHOR(S): James K. Starling

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey, CA 94942-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES: The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. I.R.B. Protocol number __N.A.__

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited

12b. DISTRIBUTION CODE
A

13. ABSTRACT (maximum 200 words)
Search and Target Acquisition (STA) in military simulations is the process of first identifying targets in a particular setting, then determining the probability of detection. This study will focus on the search aspect in STA, particularly with unaided vision. Current algorithms in combat models use an antiquated windshield wiper search pattern when conducting search. The studies used to determine these patterns used aided vision, such as binoculars or night vision devices. Very little research has been conducted for unaided vision and particularly not in urban environments. This study will use a data set taken from an earlier study in Fort Benning, GA, which captured the fixation points of 27 participants in simulated urban environments. This study achieved strong results showing that search is driven by salient scene information and is not random, using a series of nonparametric tests. The proposed algorithm, using points of interest (POIs) for the salient scene information, showed promising results for predicting the initial direction of search from the empirical data. However, the best results were realized when breaking the field of regard (FOR) into a small number of fields of view (FOVs).

14. SUBJECT TERMS
Search and Target Acquisition (STA), unaided human search, points of interest (POI), ACQUIRE, Combat XXI

15. NUMBER OF PAGES
83

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UU

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

PRIORITIZING UNAIDED HUMAN SEARCH IN MILITARY SIMULATIONS

James K. Starling
Captain, United States Army
B.S. Mathematics, United States Military Academy, West Point, NY, 2002
M.A. Leadership and Management, Webster University, St. Louis, MO, 2007

Submitted in partial fulfillment of the
requirements for the degree of

**MASTER OF SCIENCE IN
APPLIED MATHEMATICS**

from the

**NAVAL POSTGRADUATE SCHOOL
June 2011**

Author:

James Kendall Starling

Approved by:

Carlos F. Borges
Thesis Advisor

Paul F. Evangelista
Second Reader

Approved by:

Carlos F. Borges
Chair, Department of Applied Mathematics

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Search and Target Acquisition (STA) in military simulations is the process of first identifying targets in a particular setting, then determining the probability of detection. This study will focus on the search aspect in STA, particularly with unaided vision. Current algorithms in combat models use an antiquated windshield wiper search pattern when conducting search. The studies used to determine these patterns used aided vision, such as binoculars or night vision devices. Very little research has been conducted for unaided vision and particularly not in urban environments. This study will use a data set taken from an earlier study in Fort Benning, GA, which captured the fixation points of 27 participants in simulated urban environments. This study achieved strong results showing that search is driven by salient scene information and is not random, using a series of nonparametric tests. The proposed algorithm, using points of interest (POIs) for the salient scene information, showed promising results for predicting the initial direction of search from the empirical data. However, the best results were realized when breaking the field of regard (FOR) into a small number of fields of view (FOVs).

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGEMENTS

I would like to thank the Lord for the strength to face the challenges of each day. I am truly blessed and you have given me more than I deserve.

I would also like to thank my advisor, Dr. Carlos Borges, for his input along my journey in this research. You always helped me to see the bigger picture but always helped focus me in on ways that I could contribute to this area of research. I would also like to thank my second reader, MAJ Paul Evangelista, who provided a great path to direct my study. Your feedback in the initial stages was invaluable; I appreciate your patience as you helped me understand the ins and outs of the previous study and the nonparametric tools I used in this thesis. Thanks are also in order to the Combat XXI "resident expert," Dr. Imre Balogh. You have helped me to understand what is exactly going on "under the hood" in the simulations I am trying to improve. I hope I have provided you with results that will improve future simulation models.

Thanks to the Applied Mathematics and Operations Research professors that I have had the pleasure to interact with over the past two years. I have enjoyed the professionalism and level of instruction across a wide range of classes. Please continue to bring your best to future students so that we may go out and contribute to the fight. Thanks also to the post-docs here in the math department, especially to Drs. César Aguilar and Shiva Gopalakrishnan. It was interesting to see your areas of research and to bounce ideas off of you. I appreciate greatly the time you took to answer my questions in a variety of areas.

To my fellow students, I also say thanks for the day-to-day interactions and for being a sounding board for ideas in my research and classes. I look forward to instructing cadets together over the next few years.

Finally, to my wife of over eight years, Joni, I can't thank you enough for putting up with me and taking care of our three precious daughters. You handle things with a grace that I wish I could possess. Thanks for all that you do. I look forward to spending the rest of my life with you by my side.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

Search and Target Acquisition (STA) in military simulations is the process of first identifying targets in a particular setting, then determining the probability of detection. Much work has been done previously on determining the target acquisition process and search when using forward looking infrared (FLIR), binoculars, or a night vision device. This type of search is referred to as aided search. Little work has been done with creating a search algorithm that mimics human search behavior with unaided search, specifically in an urban environment. The most widely used search algorithm, known as ACQUIRE, uses a crude left to right and right to left search queue of a field of regard (FOR) that is divided into smaller fields of view (FOVs) for a computer entity.

As an entity moves through the simulated environment, each FOV is given an equally weighted, albeit stochastic, amount of time to search. These FOVs are queued in a "windshield wiper" fashion that is independent of the entity's surroundings. For example, an FOV that has no significant scene features and where it is physically impossible for a target to be located, such as a blank wall, receives just as much time as another FOV with many scene features and many locations where a enemy target can feasibly be located.

This work is a continuation of work previously done by Evangelista, Darken, and Jungkunz [1]. The data set used will be referred to as the tier II data. It was collected during an experiment conducted by TRADOC Analysis Center, Monterey (TRAC-MTY), in April, 2009, at Fort Benning, GA. The tier II data consists of fixation locations from 27 participants, each viewing 16 scenes in a simulated urban setting. Chapter III will explain how the data was collected in more detail.

A. REQUIREMENT FOR MODELING

Live experiments to test Soldier systems are costly and time consuming. Gathering the necessary data from live participants on a large scale can be cumbersome and may not

contain the necessary details for individual systems. Replicating the experiment in various environmental conditions, such as extreme temperatures, varying degrees of visibility, or varying weather conditions, is nearly impossible to achieve. Modeling will allow the user to adjust these environmental conditions while keeping the same parameters of the system being evaluated. It is imperative that the parameters of the system are accurately portrayed by the model. This can be accomplished by evaluating the system in isolated environmental conditions and then inferring the performance of the system when there is a mixture of those conditions.

Studying the performance of the human eye can help current models by improving the cognition of a simulated entity. "In the past, significant defense acquisitions generally focused on improving physical system behavior... Many of today's defense acquisitions focus on information based improvements" [2]. Many of the physical systems acquired for military use now are focused on improving the situational awareness (SA) of the individual Soldier, and thus their performance in combat. Improving the representation of human vision in combat models improves the representation of how we acquire information from our environment. This, in turn, will improve the representation of cognition.

Additionally, Vaughan [3] cites four main goals in modeling STA. These are: 1) better Soldier training, 2) reduced fratricide, 3) improved sensor systems, and 4) more efficient camouflage, concealment, and deception (CCD) technology. Since this study is investigating the performance of unaided vision, it can indirectly improve all of these with the exception of 3). Some applications of these goals will be discussed in Chapter VII, specifically the goal of better Soldier training.

B. OBJECTIVES OF RESEARCH

This study has two main objectives:

1. *Utilize the fixations results from the tier II experiment to determine whether or not salient scene information indicates where a human subject fixates.* This will be examined by using nonparametric rank and goodness of fit tests to determine whether fix-

ations occur uniformly or are directed by scene features. We will first conduct these tests using a uniform two-dimensional grid of the scenes, and then along the columns of the two-dimensional grid to facilitate implementation in the current framework of Combat XXI. An important step in this objective is to remove target information from the scenes to remove the effects of target presence on fixations in each scene. The three tests used to confirm this objective will be outlined in Chapters III, IV, and V.

2. *Using empirical data, develop a probability mapping to dictate how a computer entity should conduct search in an urban setting.* This mapping will be used to give order to the entity's search patterns within a model. The aim is to prioritize the search queue with the FOVs that have the greatest amount of salient scene information and have this model match the empirical data. The assumption is that humans prioritize search so that they spend the most time interrogating areas where possible threats are most likely to be, thus a computer entity will properly represent the time to detection by modeling the human behavior. This will be accomplished by weighting FOVs within a FOR that have a greater number of salient objects. Ultimately, the goal of this work is to give a more accurate representation of human search in a military model.

C. ASSUMPTIONS

Since this work is a continuation of previous work, we will use the same assumptions that were used in the previous experiments [1]. These include:

1. Eye fixations represent areas of search.

2. Target scenes represented enough of the visual field to exercise realistic target search (the scenes covered a visual field of 71 degrees).

3. Eye velocity less than 12.5 degrees per second indicates a fixation.

4. The median frame of a fixation adequately indicates the center of the fixation.

5. The urban scenes presented to subjects represented realistic urban combat target scenes.

6. Features in this study, specifically the coefficient of variance, generalize to mixed and other environments (e.g., wooded, desert).

After exploring the data further and creating a model, these additional assumptions were included:

1. Excluding fixations that are within a 100 pixel radius around targets will suffice to effectively remove target information in the scenes.

2. Aggregating the data along vertical columns will facilitate implementing the model in current simulations such as ACQUIRE and Combat XXI.

3. Points of interest (POIs) can be used to guide search in simulations.

D. THESIS ORGANIZATION

The second chapter of this thesis will discuss the psychological background for this research as well as give a general overview of the search algorithm used in many combat models. This chapter will also serve as a literature review. Chapter III will outline the methods and equipment used in the previous study to quantify the fixation data. It will also discuss the results of the complete data set and compare those to the previous study.

Chapters IV and V will outline two methods used to affirm the first objective. The first method, discussed in Chapter IV, removes fixations located in proximity to targets and uses a nonparametric test to affirm that search is driven by salient scene information. The second method, discussed in Chapter V, aggregates the fixation data along columns to allow easier implementation into the ACQUIRE algorithm. This method uses a different nonparametric test to also affirm the first objective.

Chapter VI proposes a model to change the queuing order in the ACQUIRE algorithm. This chapter will attack the second objective of this research, using the information gleaned from the first objective. The model compares saliency mappings, using a two variable Gaussian function, calculated from POIs and the fixation data. Some of the findings prompt us to ask the opposite question in future research; instead of "where should we direct the gaze of a computer entity?", we should instead ask, "where should we not direct the gaze of a computer entity?" The final chapter reiterates the findings of this thesis as well as discusses promising future areas of research.

THIS PAGE INTENTIONALLY LEFT BLANK

II. SEARCH THEORY

A. HUMAN VISION AND SEARCH

The way we interact with the world around us is primarily driven by what we see. When we operate our motor vehicles, we make decisions based on what we perceive other vehicles to be doing. When we participate in sports, we react to what the other players are doing as well as the location of a ball. Soldiers on a patrol plan, react, and make decisions based on how they perceive an enemy's actions or locations. Many of these decisions are made in fractions of a second as our environment changes around us.

One major limitation with human vision is that we can only "attend to a very limited number of features and events at any one time" [4]. Because of this limitation, we examine our environment in serial fashion where we break up our environment into "smaller, localized analysis tasks" [5]. We then prioritize these tasks and begin interrogating them in a mix of decreasing importance and some relationship to the proximity of prior fixations.

It is generally accepted that human search can be broken up into two stages [5, 6, 7, 8]. The first is a preattentive stage, where the bottom up features draw the eye to certain aspects of a scene. This can be likened to the idea that if we catch movement to the side of our vision, such as a thrown baseball, we instinctively react by crouching and simultaneously try to identify the path of the baseball to ensure that we will not be hit by it. The second stage is an attentive stage, where the observer performs a serial sequence of inspecting possible locations of the desired target. An example of this type of behavior is how we search for our loved ones while waiting at the exit gate of an airport. We systematically search the other passengers as they exit, ignoring those that do not fit the general outline of our target, until we do in fact find them.

The preattentive stage happens at the onset of a scene being presented to an observer. During this stage, the brain identifies prominent features and objects that are placed in a set of maps that the brain references in the later stage of search [9]. The brain does this

almost instantaneously and is based on a number of features of the scene. This first stage is accomplished when light received by the eye is "converted into a coded description of lines, spots or edges and their locations, orientations and colors" [9]. This stage is done automatically and does not require overt attention. The details and identification of these general objects is left to later stages. Torralba [10] examined the idea of more general image properties and their affect on search. He states that "early scene interpretation may be influenced by global image properties that are computed that do not require selective visual attention." His work shows that low level features can "reliably predict the semantic classes of real world scenes." This stage breaks down the scene into objects which are then interrogated further in the next stage. The amount if time taken to conduct this initial recognition stage is quick; often only "150 msec after image onset" [10].

The result from this preattentive stage is the idea of "pop-out," as mentioned in much of the literature [6, 9, 11, 12]. "Pop-out" is the idea that certain features in scenes draw our attention more than other aspects. These features are identified in the preattentive stage, as mentioned above, which are "created by the combination or arrangements of components" [6]. Doll and Home [11] discussed the idea of being able to train observers to pick out certain features quickly through training. One example they give is that of trained military personnel who can "immediately pick out targets in cluttered scenes that novice observers must search for painstakingly" [11]. The idea here is that simulation training, with real-world targets, can reduce the amount of time an observer takes to search and identify a target.

The second attentive stage is volitional and is based on top down cues [5]. This stage is much slower and is dictated by the important aspects identified by the preattentive stage. For Soldiers, this stage is usually affected by mission requirements, available intelligence, and previous experience in their current setting. The attentive stage is also slower since it is a serial process defined as a "visual and motor interaction with the world characterized by the convergence of gaze toward a target" [8]. The attentive stage attempts to determine more information about the features discovered in the preattentive stage.

This second stage of search can also be broken up into two distinct aspects. The first aspect consists of "high velocity movements that serve to move the fovea from one fixation location to another" [4]. These movements are referred to as saccades or saccadic movements. It is estimated that saccades can happen as often as "three to five" per second and are "our most observable behaviors" [8]. The second aspect is where the eye movements are of a much slower velocity; these are labeled as fixations.

It is also important to discuss the idea of attention. Attention is another aspect of vision that is not directly observable, but dictates how we perform saccades and fixations. Changes in our attention do not always correlate with an observable change in fixation [8, 10]. The "oculomotor plant" consists of the muscles around the eye and all of the "neural machinery in the brain stem that controls them" [13]. External observers can see the effect of a change in our attention as we change our gaze. Eye movements are thought to be dictated by movements of attention to a "target location before actual movement is deployed" [10]. Since the changes in attention are not easily measured, but changes in the viewing angle of the eye are, we assume that the information from saccades and fixations can be used to approximate how attention is directed. Zelinsky [8] states that "oculomotor scanning, and not purely covert shifts of attention, may be the more natural search behavior during a free-viewing task."

The data used for this study assumes that eye movements, i.e., saccades and fixations, are driven by changes in attention. Saccades in this study are inferred to be areas of attention from a directly observable sharp change in eye angle using a device attached to the participants' heads. Fixations are inferred where the change in eye angle is a smaller rate. Chapter III will explain the this process in greater detail.

B. APPLICATION TO MODELING

The oculomotor system is very complex and many of the details are still not understood. Much of the body of research, however, agrees on one point: that, as Zelinksy [8] puts it, "search is guided." With this in mind, any model of vision should include some

aspect of driving the search in its algorithm. However, if an algorithm is too simplistic, it can omit many factors that will accurately portray the performance of the human visual system. Specifically, military modeling is criticized as "emphasiz[ing] only a part of the neural 'machinery' involved" in human search [11]. If the model is too complex it can lead to being too unwieldy for plausible implementation into a simulation. The answer for a proper model lies somewhere in the middle; it must be specific enough to not affect the physical performance of the system, but it also must take into account multiple factors to accurately model the human visual system. Vaughan [3] reiterates this point by stating, "there is undoubtedly a benefit to models that take more than a single factor into account." He identifies movement, contrast, scene clutter, target velocity, and range as some additional factors that must be taken into account when developing a model.

Zelinsky [8] discusses the downfall making a general purpose algorithm that allows for user specified parameters. He instead developed a Target Acquisition Model (TAM) that focuses on a handful of principles to create. One of these principles is to "retina-transform the visual scene for every eye fixation." This would require a rendering of each scene as it runs the model. The computational cost of this would be very expensive and therefore not realistic for use in ACQUIRE. Another principle he mentions is to "represent visual information using high dimensional vectors of simple features." This study hopes to capitalize on this idea by tagging scene information with different values to guide search. The overhead in ACQUIRE for representing these features would be minimal and would fit into the current algorithm nearly seamlessly.

Human search models must also implement the preattentive and attentive stages into their algorithms to properly represent human vision. Identifying the global scene characteristics and creating a map of these objects and features will suffice for tackling the preattentive stage [8, 10]. Possible methods to achieve a mapping of these features include identifying possible hiding spots [14], or limiting the scene to horizontal regions based on possible target locations [10]. Gaussian functions will be utilized in our proposed model and discussed further in Chapter VI. The second attentive stage can be modeled using the

information extracted from the first stage. A saliency mapping of the important features can be computed and used to direct "the deployment of attention and first eye movements toward likely locations of target objects" [10]. As an entity moves through a simulated environment, the first preattentive stage will need to be employed multiple times as new features become observable. This can be done by discretizing the virtual environment and precomputing the saliency maps for each region. Evangelista, Ruck, Balogh, and Darken [2] present this idea of preprocessing information on line of sight (LOS) information. The LOS information becomes a "characteristic of the terrain, stored in a database for future lookup during runtime." The benefits of preprocessing LOS, and thus scene information, cannot be underestimated; it can allow faster run times of the actual ACQUIRE algorithm and also help guide search.

Many studies can now take advantage of simple and complex environments to measure how human observers react to differing levels of stimuli. The setting in which Soldiers find themselves in current military operations and in urban environments, often in third-world countries, are complex. These environments contain many corners, roof lines, windows, doors, rubble, trash, etc., that must be scanned while simultaneously performing their mission. Much of the body of knowledge about this subject used natural scenes, i.e., real photographs, which can be used for greater study of the effects of complex environments and better representation of STA [5, 8, 10]. The scenery used for this study was complex, albeit computer generated. It is important to note that Combat XXI uses a very simple representation of the environment and cannot match the complexity of real-world scenes without sacrificing performance.

This study focuses on the search aspect of STA and does not explore the actual acquiring of targets; that is done independently of search in Combat XXI. The proposed model also does not attempt to directly address the preattentive stage, but rather focuses on the location of fixations in the scenes in a general sense to guide search. An archetypical observer "will fixate the image locations that have the highest probability of containing the target object given the available information" [10]. Keeping this in mind, the model will

use salient scene information to develop a saliency map that will be used to determine the order in which a computer entity interrogates FOVs in a specified FOR.

C. ACQUIRE—TIME LIMITED SEARCH

Combat XXI currently uses ACQUIRE as its search algorithm. ACQUIRE uses a simple "windshield wiper" method that sweeps across a specified field of regard (FOR) that is broken into smaller field of views (FOVs). The FOVs are adjacent and nonoverlapping. For each FOV, the simulation determines two sets of times from different distributions: the first is the amount of time to detect any targets, if present, and the second is the amount of time to interrogate the FOV. If there are multiple targets present, the simulation will calculate a time to detect for each target. The time to interrogate a FOV is also called the empty field of view time (EFOV). If any of the times to detect a target is less than the EFOV time, then the entity is determined to have detected the target. If, on the other hand, the time to interrogate the FOV is less than the time to detect the target, the entity will not detect the target and will move to the next adjacent FOV [15]. If there are multiple targets in a FOV, any additional detections are determined by adding an additional second to the search time and then compared to the other targets' detection times. If the other detection times are greater than this new time, then the additional targets are considered to be detected as well [15].

This search model that is described above is called the "Time Limited Search" (TLS). The time to detection is determined by using three different formulas based on the surrounding type of the target:

$$\begin{aligned}
\text{Rural FOR:} \quad & \text{time}_{det} = (3.5 - 2.5 \cdot P_{\infty}) \ln(1 - RN(0,1)) \\
\text{Urban FOR (human target):} \quad & \text{time}_{det} = (5.57 - 3.89 \cdot P_{\infty}) \ln(1 - RN(0,1)) \\
\text{Urban FOR (other targets):} \quad & \text{time}_{det} = (3.5 - 2.5 \cdot P_{\infty}) \ln(1 - RN(0,1))
\end{aligned}$$

where P_{∞} is the probability of detection of a specific target type given an infinite amount of time and RN(0,1) is a random number uniformly distributed from 0 to 1.

The calculation of the EFOV is also broken down into four contextually driven formulas:

$$\text{Urban FOR (human target):} \quad EFOV = -2.59 \cdot \ln(1 - RN(0,1)) + 1$$

$$\text{Urban FOR (veh/human target):} \quad EFOV = \sqrt{\frac{-0.69392}{\ln(RN(0,1)) - 0.000771}}$$

$$\text{Rural FOR (mod clutter):} \quad EFOV = \sqrt{\frac{-11.1}{\ln(RN(0,1)) - 0.06}}$$

$$\text{Rural FOR (high clutter):} \quad EFOV = \sqrt{\frac{-20.4}{\ln(RN(0,1)) - 0.12}}$$

The flow chart of the general TLS methodology is shown in Figure 1 [16].

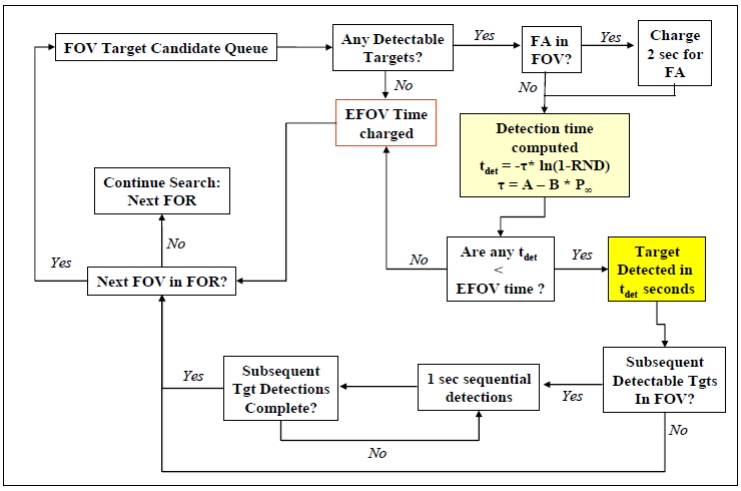

Figure 1: TLS FOR methodology flow chart. [From [16], Figure 1, p. 3]

The first issue with the current TLS and Urban FOR model is that they do not model the human eye as a sensor. In the studies conducted by Harrington [15], Jones and Lai [16], and Grove [17], all of the images used in the TLS and the Urban FOR studies were 1st and 2nd generation Forward Looking Infra-Red (FLIR) images. These sensors usually have two FOV settings: wide and narrow. It is our hypothesis that humans conduct search in a different manner when searching with unaided vision versus aided vision.

The second issue with the current model is that the FOV Target Candidate Queue, shown at the upper left side of Figure 1, follows a simple left to right and right to left search pattern. It is our hypothesis that unaided search is a guided process, dictated by both bottom-up and top-down aspects of the scene. The current candidate queue assumes

that search is random, but research has shown otherwise. Doll and Home [11] have found that search is dictated by "objects that are most conspicuous" and that "clutter drives visual search." The second objective of this study is to change the simple left to right queuing by determining which FOVs in a FOR are most important and should be interrogated first.

III. FIXATIONS AS REPRESENTATIONS OF ATTENTION

Since human vision is a very complex process, we will use eye fixations as a determination of attention and areas of search. Zelinsky [8] stated that "eye movements are directly observable; movements of attention are not." Dorr [4] separated eye movements into two distinct classifications: saccades and the movements that are made between saccades. Although he describes multiple ways to classify the movements between fixations, we will simply classify these as fixations. As stated in the assumptions section above, we will use eye fixations as representations of areas of search, or attention. This section will describe the method used to capture the fixation locations and discuss the methods used to quantify unaided vision. The results from this section will be used to provide evidence towards confirming the first objective of this study, that search is driven by salient scene information.

A. TIER II FIXATION DATA

The tier II experiment was planned, executed and led by TRAC-Monterey. It was conducted at the Maneuver Battle Lab in Fort Benning, GA in April 2009. The experiment was conducted over nine days where eye tracking data was collected from 27 infantry Soldiers, the majority of whom had combat experience in current operations. Each participant used a full-sized rifle that was instrumented with the simulation to capture shots fired and hits for 16 different urban scenes. The number of targets in the scenes ranged from zero to five targets, with a total of 39 targets for all 16 scenes. Table 1 shows the number of scenes with their respective numbers of targets.

During the experiment, video images were captured using the Mobile Eye tracking device manufactured by the Applied Science Laboratories. This equipment recorded videos for each participant in $1/30^{ths}$ second frames for each scene with the associated x and y

Table 1: Number of scenes containing targets

Targets	Number of Scenes
0	2
1	2
2	4
3	4
4	3
5	1

coordinates of the foveal point, relative to the equipment's field of view, and thus to the observer's field of view as well. Figure 2 shows two example frames from these videos.

Figure 2: Two example fixation frames

Evangelista, Darken and Jungkunz [1] spliced these video files into smaller files where they calculated eye movements to be less than 12.5 degrees per second. This step reduced the number of usable fixation data from the 27 to 25 total participants. This was due to the fact that two of the participants moved their heads and eyes too erratically to properly classify fixations. With these smaller video files, they captured the median frame of fixation to approximate the fixation location. The experiment allowed free range of motion for the participants' heads, which necessitated the need for an approximation to the fixation location. A manual process was utilized to capture the coordinates of the fixation from the median video frames. The process was conducted by calibrating a keyboard emulator to capture the x and y coordinates of the scenes, which were in jpeg format. They then approximated the fixation location by first locating the cross hairs on the median frame and

then clicking on the scene. The emulator captured the coordinates and the fixation number in a text file, where the fixation number was a counter that started at 0 and increased by 1 for each fixation. These files were later appended to include the participant number, as well as the starting and ending frame numbers.

Prior to this study, approximately two-thirds of the fixations were captured in the data set. The remaining one-third had to be collected using the same method as described above. This study also performed a sanitization of the data in order to remove fixations that occurred before the 20 second time period or fixations that were indeterminable. After removing these fixations, we had a data set with 9179 fixations. Figure 3 shows examples of a fixation that occurred before the time period and an indeterminable fixation.

Figure 3: Examples of removed fixations. **Left plot:** Fixation that occurred before start of experiment. **Right plot:** Indeterminable fixation.

From the file with the full set of sanitized fixations, we can then graphically show where these fixations occurred using the coordinates and the scene number. Figure 4 shows an example of fixation data from one scene. For this particular scene there are two targets highlighted in red and fixations are represented by the blue dots.

B. EYE TRACKING EQUIPMENT

The Tier II data used the Mobile Eye tracking device manufactured by the Applied Science Laboratories. The Mobile Eye tracking device consists of the spectacle mounted unit (SMU), spectacles, digital video cassette recorder (DVCR), and a recorder mounted

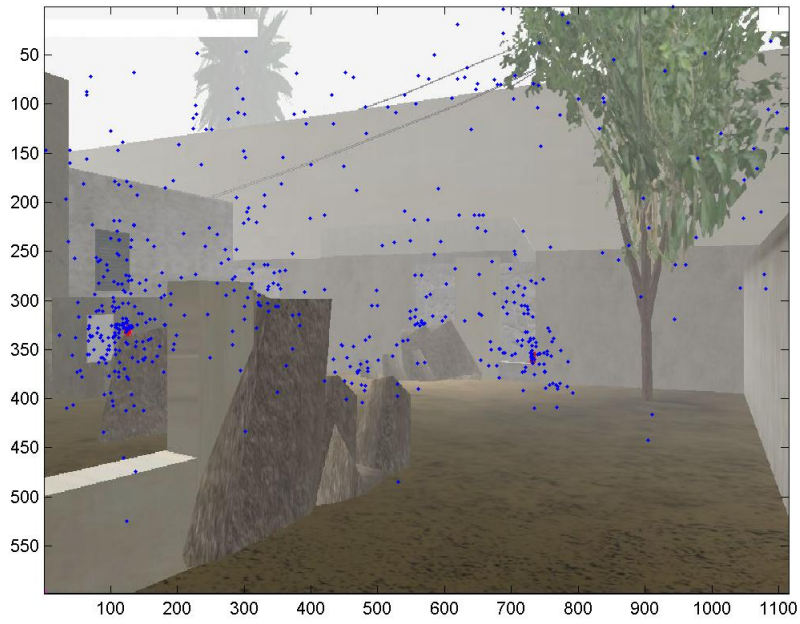

Figure 4: Example scene with fixations (represented by blue dots)

unit (RMU). The SMU is attached to the spectacles and worn as normal spectacles are worn. The weight of the spectacles and the SMU was insignificant, and the participants, "stated that the system felt much like the safety glasses worn during combat operations" [1].

The SMU consists of two cameras; one camera tracks the movement of the eye and a second camera records the scene information. There is also a set of three near infra-red lights on the SMU that is projected into the observer's eye. The eye camera records the "corneal reflection," which is the "relationship between two eye features, the pupil and a reflection from the cornea" [18]. Using the scene camera in conjunction with the corneal reflection data, the Mobile Eye tracker can determine the coordinates of the eye gaze, with respect to the scene camera's view, or the participants' heads. The tracking can be shown via a circle, small crosshair, or large crosshair. All necessary calibration was conducted for each participant prior to being exposed to any scenes in the experiment. Examples of some of the video frame captures, with large crosshairs, are shown above in Figures 2 and 3.

C. SCENE DISCRETIZATION

In order to translate the fixations into a form that can be more readily analyzed, each scene was initially discretized into a 30 $\times$ 30 mesh along the horizontal and vertical axes. Each discretized cell in this mesh was assigned a vector of independent variables. These independent variables included the two-dimensional distance to doors, windows, moving target locations, target locations, and audible cues [1]. The distances were computed using the Euclidean distance from the fixation point to the center of the item of interest, such as the center of a door or target. Another dimensionless, independent variable was used to determine the change in depth at a particular cell. This is called the coefficient of variance for line of sight (LOS). The previous study used the LOS values, or the distances from the observer to center of a cell, as calculated by the virtual environment simulation used in Fort Benning, GA. To calculate the coefficient of variance for each cell, they used the standard deviation of the cell and its eight neighboring cells divided by the sample mean of these same nine cells [1]:

$$cv_{ij} = \frac{\sigma_{ij}}{\mu_{ij}}$$

Where μ_{ij} and σ_{ij}^2 are defined as follows:

$$\mu_{ij} = \frac{\sum_{a=i-1}^{i+1} \sum_{b=j-1}^{j+1} LOS_{ab}}{9}$$
$$\sigma_{ij}^2 = \frac{\sum_{a=i-1}^{i+1} \sum_{b=j-1}^{j+1} (LOS_{ab} - LOS_{ij})^2}{n-1}$$

The outer cells of the mesh had to be removed in order to calculate the coefficient of variance. Thus, the actual mesh used in the study was reduced to a 28 $\times$ 28 mesh. This gives a data set of 12,544 discrete cells when including all 16 scenes. Table 2 shows the definitions of these variables.

Table 2 also shows dependent variables f_{ij}, n_{ij}, and t_{ij}. To determine the values of f_{ij} we do the following: for each fixation, we determine which discrete cell is closest

to the fixation point and then make the value of that discrete cell equal to 1. If a discrete cell does not have a fixation near it, it will be given the value of 0. For each discrete cell that has been marked with a 1, we also increase n_{ij} by 1, and increase t_{ij} by the length of the fixation. It is noted that the original study only included the dependent variable f_{ij}; the additional dependent variables, n_{ij} and t_{ij}, were added in this concurrent study. This was done to allow further study into the number of times a cell was fixated on and also to study the amount of time spent interrogating a cell.

Table 2: Definition of variables for discrete cells

Independent Variable	**Definition**
w_{ij}	Distance in pixels from discrete area (i, j) to nearest window
d_{ij}	Distance in pixels from discrete area (i, j) to nearest door
m_{ij}	Distance in pixels from discrete area (i, j) to nearest moving target
a_{ij}	Distance in pixels from discrete area (i, j) to nearest audible cue
cv_{ij}	Coefficient of variance of the LOS to (i, j).
Dependent Variable	**Definition**
f_{ij}	1 if a fixation occurred at discrete point (i,j), 0 otherwise
n_{ij}	Number of fixations at discrete point (integer)
t_{ij}	Total time spent at discrete point ($1/30^{ths}$ seconds)

D. MANN-WHITNEY *U* STATISTIC

Since the f_{ij} data is in a binary classification format, we can apply the Mann-Whitney U statistic which will give us the probability that a random positive instance is ranked higher than a random negative instance [19]. This statistic is computed by ranking the f_{ij} variables; each discrete cell is assigned a rank, R_i, as determined by the value of the independent variables, i.e., w_{ij} or d_{ij}. Lower ranks are assigned to the mesh points with shorter distances to a specified object, such as a door or window, since we are placing more importance on cells that are closer to these objects. Since the coefficient of variance

measures changes in depth, we assign lower ranks to the cells that have a greater coefficient of variance. The following equation shows how to calculate the U statistic:

$$W_r = \sum_{i=1}^{b} R_i$$
$$W_{YX} = W_r - \frac{1}{2} b \cdot (b+1)$$
$$U = \frac{W_{YX}}{p \cdot b}$$

where p is the number of fixations, b is the number of nonfixations, and R_i is the rank of the ith fixation. It is important to note that the ranks being summed when calculating W_r are only those from the benign instances. The resulting values for U will range between 0 and 1. If the cells that were fixated on dominate the lowest ranks and every fixated cell is ranked higher than nonfixated cells, the value of the U statistic will be one. If the cells that were not fixated dominate all of the lowest ranks, then the U statistic value will be zero. The previous study found that ranking the cells according to only one of the independent variables yielded values of the U statistic between 0.70 and 0.59 [1]. This study found similar values when ranking the cells using only one feature.

In order to achieve greater performance, it is possible to use the ranks of the independent variables and then use an aggregator of these ranks. Some possible methods are to use the minimum, maximum, or mean of these rankings. Evangelista, Darken and Jungkunz [1] found that using all five independent variables by way of using a minimum aggregator yielded significant results over using just one variable ($U = 0.78$). They also discovered that since the moving targets and audible cues only occurred in approximately half of the scenes, by removing them from the calculation, they could achieve even better results ($U = 0.79$). For this, they defined $rank_{ij} = min[rank(w_{ij}), rank(d_{ij}), rank(cv_{ij})$. The rest of this study will use the min aggregator of these three independent variables when dealing with the fixation data.

1. Mann-Whitney *U* Statistic Example

To clarify how this statistic was used in this study and the previous study, we will show a small example. We will assume there are 10 cells that we are examining, labeled a through j. With these 10 cells, we will create three independent variables derived from random normal, poisson, and weibull distributions. These will be labeled x_1, x_2, x_3, respectively. We first rank each of the variables according to their own kind. With these rankings we then take the minimum of the three rankings. Table 3 shows the random variables' values, their rankings, and their aggregated rankings using the min function.

Table 3: Example of three independent variables from random distributions

cell	x_1	rank(x_1)	x_2	rank(x_2)	x_3	rank(x_3)	fixation	agg. rank (min)
a	14.52	4	96	6	11.04	10	1	4
b	20.68	5	90	3	6.90	4	0	3
c	25.12	6	94	5	2.36	1	0	1
d	31.43	9	112	10	10.61	9	1	9
e	11.06	1	103	8	7.64	5	1	1
f	14.28	3	87	2	7.99	6	0	2
g	12.79	2	85	1	9.68	8	1	1
h	28.47	7	97	7	9.13	7	1	7
i	35.72	10	91	4	4.78	3	1	3
j	30.55	8	109	9	4.67	2	0	2

The next step is to order the rows by their minimum rank and assign an absolute ranking, as shown in Table 4. It is important to note that with the minimum aggregator there are repeated values, or ties, of the rank values. With small sample sizes, this will cause some fluctuation in the U statistic. Lehmann [20] recommends using midranks, or the mean of the ranks, where ties occur. In our study, however, our sample size is large enough to assume the central limit theorem and the number of ties is insignificant. For this example we find that $p = 6$ and $b = 4$. To calculate W_r, we simply sum the absolute ranks where the fixation column is 0. This yields $W_r = 1 + 4 + 5 + 6 = 16$. We then calculate $W_{XY} = 16 - \frac{1}{2}(4)(4 + 1) = 6$, and our U statistic $= \frac{6}{6 \cdot 4} = 0.25$.

Table 4: Assigning absolute rank after sorting according to aggregate rank

cell	fixation	agg. rank (min)	abs. rank
c	0	1	1
e	1	1	2
g	1	1	3
f	0	2	4
j	0	2	5
b	0	3	6
i	1	3	7
a	1	4	8
h	1	7	9
d	1	9	10

E. RESULTS FROM THE COMPLETE DATA SET

As described above, the previous study only considered the results from approximately two-thirds of the participants. The final third of the data points were added with the keyboard emulator method and then sanitized as described above giving the final data set of 9179 fixations. Using the min aggregator, as did Evangelista, Darken, and Jungkunz [1], we discovered a U statistic of 0.7802, compared to 0.79 in the first study. Although this shows a slight drop, it still shows a strong relationship between fixations and changes in depth and proximity to doors or windows. The results from the complete data set provide the first steps of confirming the first objective in this research. Chapters IV and V will explore further methods to confirm the first objective.

The full data set had a b value of 9295 and a p value of 3249, compared to b = 9280 and p = 3264. These p values initially caused alarm, since the full set of fixations showed that 3249 cells had been fixated upon, which was a decrease of 15 cells. This can be explained in that in the sanitization discarded 372 fixations. It was discovered that many of these fixations had been mapped to the 28×28 mesh in the first study.

THIS PAGE INTENTIONALLY LEFT BLANK

IV. TARGET REMOVAL

One of the issues concerning the test conducted at Fort Benning, when determining the effect of salient scene information, is the presence of targets in the urban scenes. The study from which this data originates had targets in all but two scenes, as shown in Table 1. The previous experiment was designed to measure the participants' abilities to detect and engage targets. Since the participants were engaging targets, we would expect that a majority of fixations occurred near target locations. The manner by which fixations were determined did not discriminate according to proximity to target locations; fixations were only classified for eye velocity less than 12.5 degrees per second as stated in Chapter III. An example of fixations clustered around targets can be seen below in the upper left quadrant of Figure 5. This chapter will continue to confirm the first objective of this study by showing that fixations are indeed drawn to salient scene information, even when target information has been removed from the scenes.

In order to circumvent the problem of targets drawing attention in the scenes, fixations within specified radii around a target were removed. The radii selected were 0, 100, 200, 300, 400, 500, and 1000 pixels. A radius of 0 pixels corresponds to not removing any cells and radius of 1000 pixels corresponds to selecting the scenes that did not have any targets present—only 2 scenes. Figure 5 shows a graphical example of removing various radii around a target.

When removing the fixations we must also remove the discrete cells in the mesh that are within the specified radii. Figure 6 shows a graphical example of removing the cells within various radii around a target.

The number of fixations accounted for with a radius of zero was the full number of fixations, 9179. However, we experienced a significant reduction when removing a radius of 100 pixels when number of fixations dropped to 4675. This shows that approximately 50% of the fixations occured within a 100 pixel radius around a target, which is an inordinate amount considering the pixel dimensions of the scenes. As we continued to increase

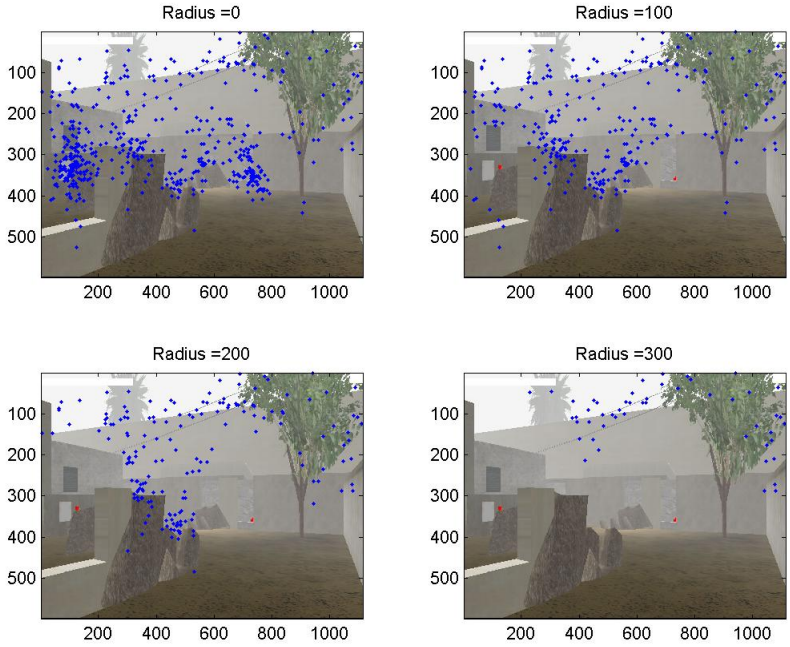

Figure 5: Example of removing fixations with various radii

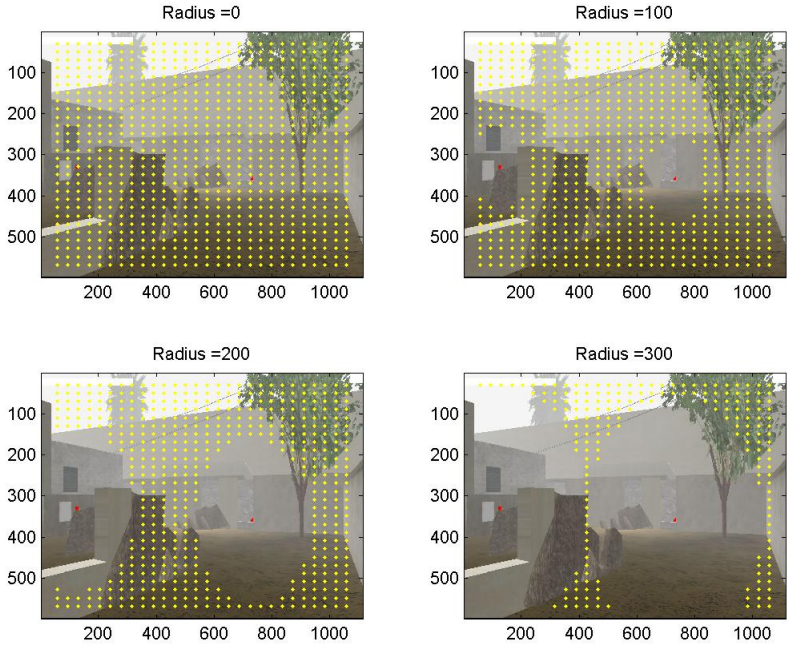

Figure 6: Example of removing discrete cells with various radii

the radius for removing fixations, the number of fixations continued to decrease, but we did not experience as sharp as a drop as we did after removing only 100 pixels. Table 5 shows the number of fixations accounted for as we increase the radius from zero to 1000.

Table 5: Number of fixations after removing various radii around target

Radius	Number of Fixations
0	9179
100	4675
200	2720
300	1790
400	1408
500	1300
1000	1246

When first removing the fixations that are within a given pixel radius and subsequently removing the discrete cells within the same radius, an issue arises where some of the fixations that are outside of the radius are mapped to a cell that is within the radius. The net result is that some of the fixations are lost due to this method. Table 6 shows the number of fixations and their percentage of total fixations that are removed due to this process for each radius. For example, with radius 100 the total number of fixations in our data will only be 4574, not the 4675 as shown above in Table 5, but this is only a small percentage of the total number of fixations in our pool. This will not prove to be an issue since removing these additional fixations takes a more conservative approach to removing fixations around targets, which is our ultimate goal with this method.

The U statistic is then calculated for each radius. Table 7 shows the results of the chosen radii values and their effect on the total number of cells, U statistic, benign instances, and positive instances for each radius. The U statistic values for all radii not equal to zero shows improvements over the zero radius U statistic. This shows that the targets' presence was drawing attention away from the other scene information such as changes in depth, as represented by the cv values, and the distances to doors and windows.

Table 6: Number of fixations removed from discrete cells removed from various radii

Radius	Fixations Removed	Percent of Total	Total Fixations
0	0	0	9179
100	101	0.0216	4574
200	68	0.0250	2652
300	12	0.0067	1778
400	30	0.0213	1378
500	4	0.0031	1296
1000	0	0	1246

One significant result to notice is that for the scenes with no targets, the U statistic is better than when there is no target information removed (0.7986 with no targets vs. 0.7802 with the full data set). These results reinforce what was discovered by the previous study that salient scene information, specifically changes in depth and the presence of windows and doors, tend to dictate where a human subject fixates even after the data has had target information removed.

Table 7: U Statistics after removing various radii around target cells

Radius	# of Cells	U statistic	benign instances	positive instances
0	12544	0.7802	9295	3249
100	10911	0.7875	8873	2138
200	7408	0.7942	6210	1198
300	4137	0.7933	3407	730
400	2419	0.8017	1905	514
500	1822	0.7999	1353	469
1000	1568	0.7986	1126	442

A. CHOOSING THE PROPER RADIUS

For the tier II experiment, the participants stood 7 feet from a 10 foot wide by 7.5 foot tall screen [1]. The scenes that were presented to the participants were 1114 x 598 pixel (width x height) images, which equates to a 71.1° viewing angle. This equates to 15.67 pixels per degree along the horizontal axis. Assuming that the human eye fixates

with a foveal point of 2°, a fixation covers approximately 31 pixels on the screen. So, for any given target, a fixation can cover a diameter of approximately 62 pixels. Accounting for other types of errors in our determination of fixation points, we will focus on using only a 100 pixel radius and the effect of the U statistic.

As shown in Table 5, using a radius of 100 pixels around a target removed approximately 50% of the fixations. Removing fixations, and the corresponding discrete points, within a 100 pixel radius around targets yields an U statistic of 0.7875, which is slightly better than when not removing any target information. Although there are greater results when removing a larger radius, we will not use those results and instead assume that removing a radius of 100 pixels will be sufficient to remove fixations influenced by a target's presence. This second test gives further evidence that the first objective of this study is indeed true. For the remainder of this study, we will assume that the 100 pixel radius around a target will suffice for removing fixation information due to the presence of targets.

THIS PAGE INTENTIONALLY LEFT BLANK

V. COLUMN AGGREGATION

A. COLUMN AGGREGATION

Currently, the ACQUIRE algorithm only populates the FOV Target Candidate Queue with the angle of the sensor in the horizontal axis but does not account for the vertical axis when conducting search. In light of this, we will aggregate the data along the vertical columns for all 16 scenes. Doing so yields 448 vectors of information and can change the unbalanced properties of the previous discretization. For example, Figure 7 shows which cells in a scene are aggregated along a column; this example indicates that the ninth column is aggregated as indicated by the red circles. This chapter will outline the third test used when confirming the first objective of this study.

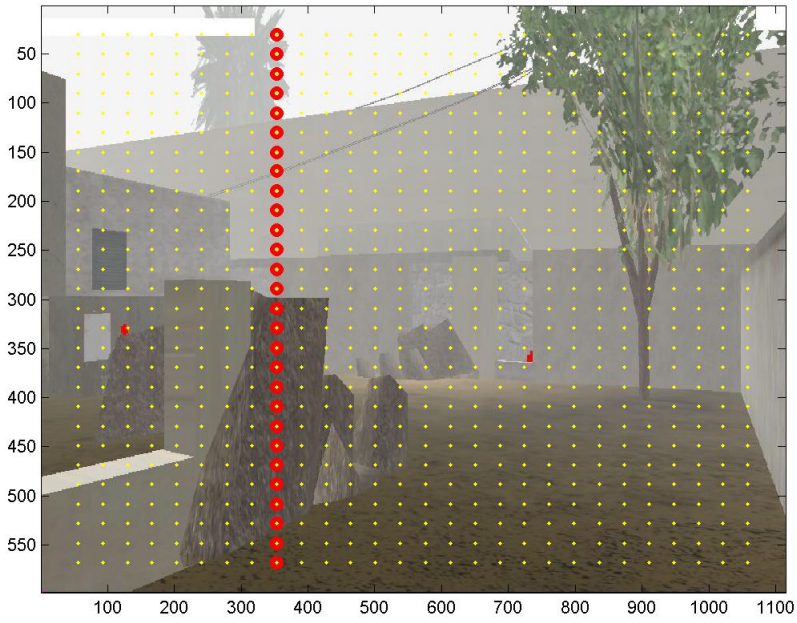

Figure 7: Aggregating column data. In this example, the ninth column is highlighted with red circles to show which discrete cell information is used for the ninth column

For each column we assign a vector, which will have the same variables listed in Table 2, but with slightly different definitions. For windows and doors, we want to represent the column by how close it is to a window or door. Thus, the distances to windows and doors, w_j and d_j, will take the minimum value along each column. Likewise, for changes in depth we want the represent each column with the biggest change in depth. Thus, the coefficient of variance, cv_j, will take the maximum value along each column. For the number of fixations and time spent at each fixation, n_j and t_j, we will take the sum along each column. Table 8 shows the variables and their definitions.

Table 8: Definition of variables for column vectors

Independent Variable	**Definition**
w_j	$\min\{w_{ij}\},\ \forall j$ (outside of radius)
d_j	$\min\{d_{ij}\},\ \forall j$ (outside of radius)
cv_j	$\max\{d_{ij}\},\ \forall j$ (outside of radius)
Dependent Variable	**Definition**
f_j	1 if a fixation occurred along column j, 0 otherwise
n_j	$\sum n_{ij},\ \forall j$ (integer)
t_j	$\sum t_{ij},\ \forall j$ ($1/30^{ths}$ seconds)

After aggregating the columns, we will utilize the Mann-Whitney U statistic with the new aggregate data for radii of 0 and 100 pixels to determine if doing so still shows a correlation between the independent variables and whether or not a fixation occurred along that column. The results for this are shown in Table 9.

Table 9: U Statistics after aggregating column information for various radii

Radius	# of Columns	U statistic	benign instances	positive instances
0	448	0.4366	6	442
100	448	0.6540	38	410

As the table shows, we get significantly poorer results for the U statistic when aggregating the information along columns than when using the 28×28 mesh. Lehmann

[20] shows that the "U [statistic] is asymptotically normal as m and n tend to infinity." Where the values of m and n here refer to the number of benign and positive instances. By aggregating the column information, we have reduced the number of benign instances along columns to 6 for a radius of 0 and to 38 for a radius of 100. Table 7 shows the number of benign cells was 9295 and the number of positive cells was 3249 with a zero radius for the 28×28 mesh where; with a 100 pixel radius the values were 8873 and 2138 for benign and positive instances, respectively. When aggregating the information along columns, it is not preferred to use the Mann-Whitney U statistic since we can no longer assume asymptotic normality. Thus, we will instead use a different test to determine the distribution of fixations across the scenes.

B. CONTINUOUS DISTRIBUTION

In order to determine the distribution of fixations across all of the scenes, we will use the number of fixations, n_j, along each column. The method used above only allowed binary values in each column; i.e., a 1 if there was a fixation a column or a 0 if there was not a fixation along that column. Using the values of n_j will allow a more continuous approach at looking at the data by allowing a greater range of values for each column.

In order to get a feel of what the distributions of the data look like, we will plot the values of three regressor variables versus the number of fixations. The distributions of the n_j versus CV, distance to doors, and distance to windows are shown in Figure 8. Also, the histograms of the n_j versus CV, distance to doors and distance to windows are also shown in Figure 9.

The distributions of the number of fixations versus the different cv values appears to be clustered around certain values; namely 1.4, 1.9, and around 2.8, but the majority of fixations tended to occur when the cv was greater. As for the distance to a door or window, the number of fixations is clearly left skewed on the histograms and the scatter plots shown above. This shows a tendency for fixations to occur closer to doors or windows, at least

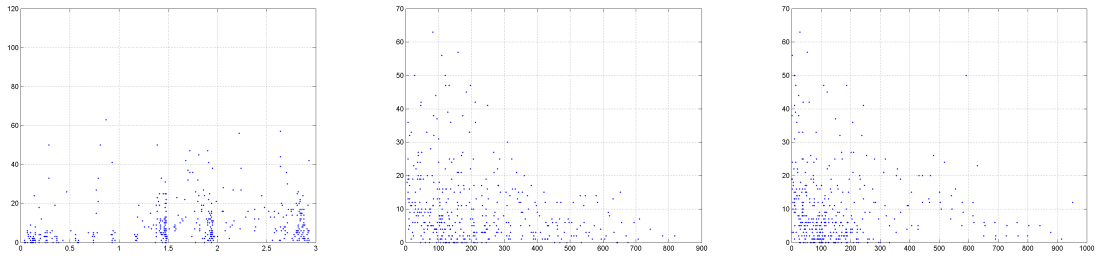

Figure 8: Distributions of independent variables versus number of fixations.
Left plot: CV values. **Center plot:** Door values. **Right plot:** Window values.

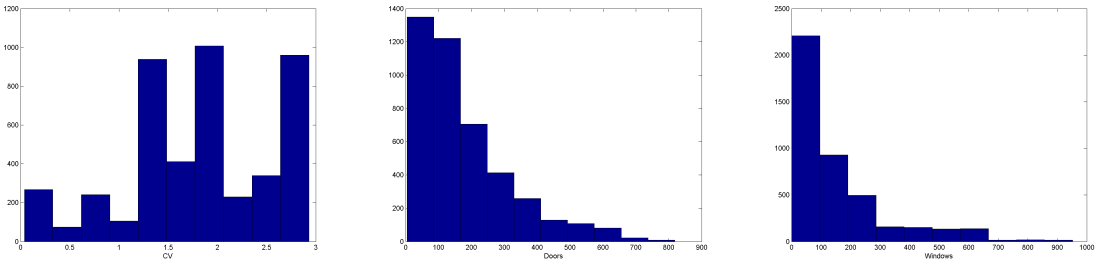

Figure 9: Histograms of independent variables versus number of fixations.
Left plot: CV values. **Center plot:** Door values. **Right plot:** Window values.

for the 16 scenes. Thus, changes in depth and the distances to doors or windows do tend to draw attention.

How then do we classify this distribution? One way is to create an empirical distribution function (EDF) from the data. If you take a sample $X_1, X_2, ...X_n$ from a population, in our case cv_j, d_j and w_j, the sample or empirical distribution function is: $F_n^*(x) = n^{-1} \sum_{j=1}^{n} \#\{x_j \leq x\}$. Thus, if we multiply F_n^* by n, we will have the number of $X_k's$ that are less than or equal to x [21]. Using the strong law of large numbers, or central limit theorem, as the number of elements goes to infinity the EDF will approach the cumulative distribution function (CDF) [22]. For our samples $n = 448$, which is the number of columns.

The EDF we wish to create here is based on the instances where fixations occurred along the columns. Each of the columns will be weighted by the number of fixations, n_j, that occurred along that column. In order to weight this properly, we will create a vector with either the CV or the distances to doors or windows for each column, repeated by the

number of fixations that occurred along that same column. Using the cv values as an example, if there are k fixations along column j, we will have the set $cv_j(1), cv_j(2), ..., cv_j(k)$. This will be repeated for each column that had at least one fixation along it. Figure 10 shows the EDFs for CV, distances to doors, and distances to windows.

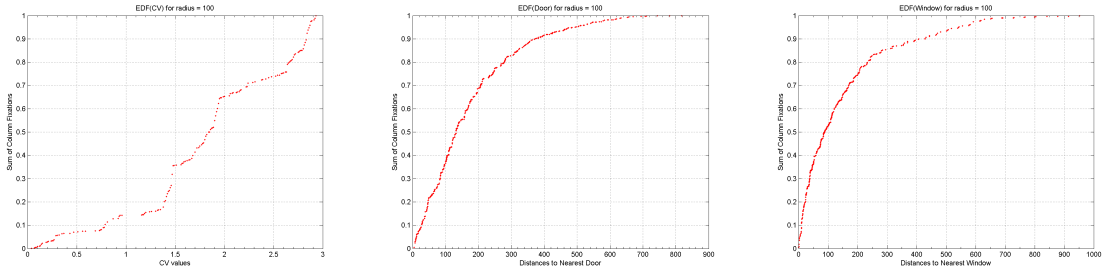

Figure 10: **Left plot:** EDF for CV values. **Center plot:** EDF for door values. **Right plot:** EDF for window values.

The EDF for the CV values shows that more fixations have occurred along the columns where the CV is greater. The EDFs for the distances to windows or doors also show that a majority of the fixations occurred near these salient scene objects. However, we must further investigate how the fixations occurred compared to a uniform distribution. The use of another nonparametric test will allow us to do so and is discussed in the next section.

C. KOLMOGOROV-SMIRNOV TWO-SAMPLE TEST

The EDFs shown above are calculated using the values of n_j weighting each column. In order to determine whether or not our data shows that the three independent variables, cv_j, d_j, and w_j, are an indicator of search, we must compare the EDFs above with EDFs from uniformly distributed values across the columns. A suitable goodness of fit test is the two sample Kolmogorov-Smirnov (KS) test.

To do this, we let $X_1, X_2, ..., X_m$ and $Y_1, Y_2, ..., Y_n$ be independent random samples from two different, continuous, distribution functions, F and G. We then let F_m^* and G_n^* be the EDF of the $X's$ and $Y's$. Note that the lengths of the two EDFs can be different. We define $D_{m,n} = \sup |F_m^*(x) - G_n^*(x)|$, $\forall x$. We will use $D_{m,n}$ to test the null hypothesis that

the fixations are uniformly distributed against the alternative hypothesis that the fixations are not uniformly distributed. We reject the null hypothesis at level α if $D_{m,n} \leq D_{m,n,\alpha}$, where $P_{H_0}\{D_{m,n} \geq D_{m,n,\alpha}\} \leq \alpha$ [21].

To calculate the significance, or p-value, for large samples we must first set $N = mn/(m+n)$. The significance is defined by:

$$\lim_{m,n\to\infty} P\{\sqrt{N}D_{m,n} \leq \lambda\} = \begin{cases} \sum_{j=-\infty}^{\infty}(-1)^j e^{-2j^2\lambda^2}, & \lambda > 0 \\ 0, & \lambda \leq 0, \end{cases}$$

Using the asymptotic distribution properties, Stephens [23] found that the upper tail of the probability for a two-sided test can be approximated by: *p-value* $= 2e^{-2\lambda^2}$, where $\lambda = D(\sqrt{N}+0.12+0.11/\sqrt{N})$. The corresponding one-sided test is approximated by half of the p-value above: $p - value = e^{-2\lambda^2}$, where λ is defined the same. The significance values shown later on in this study will use the significance values found using the KSTEST2 command in MATLAB. MATLAB approximates the two-sided test significance value using the definition of $D_{m,n}$, while using the approximation found by Stephens to estimate the one-sided test significance value [24, 25].

1. KS Example

The following is an example from Rohatgi's text [21], to illustrate the KS test by comparing the lifespans of two types of batteries. Each battery has a sample size of 6 where the lifetimes of the samples in hours are given in Table 10.

Table 10: Battery example data

Battery A:	30	30	40	40	45	55
Battery B:	40	45	50	50	55	60

Table 11 shows the possible values of battery life, x, the EDF values of batteries A and B, F^* and G^*, and the absolute value of their differences.

Table 11: Battery example results from KS test

x	$F^*(x)$	$G^*(x)$	$\lvert F^*(x) - G^*(x)\rvert$
30	2/6	0	2/6
40	4/6	1/6	3/6
45	5/6	2/6	3/6
50	5/6	4/6	1/6
55	1	5/6	1/6
60	1	1	0

The resulting value for the KS statistic from this table is $D_{6,6} = \sup |F^*(x) - G^*(x)| = 3/6 = 1/2$. A graphical example of the two battery's EDFs are shown in Figure 11. Visually inspecting the two EDFs, we can see that the greatest absolute difference between the two EDFs occurs between 40 and 45, and again between 45 and 50. This difference is the KS statistic and is equal to $1/2$. The p-value is 0.3180, and we fail to reject the null hypothesis that the two sets of data come for the same distribution at any reasonable level of significance.

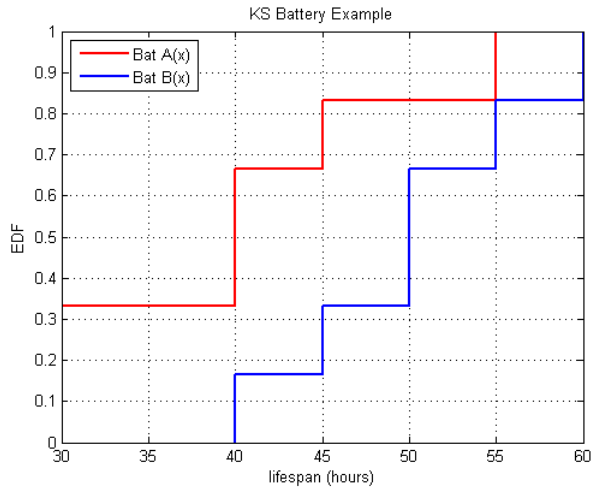

Figure 11: EDFs for two battery types

2. One-Sided KS Tests

The test described above is the two-sided KS test and is used in the example. In order to conduct a more accurate test, we should use the one sided KS test. The "greater than"

KS test is the same as the two-sided test, with the exception that you remove the absolute value function, i.e., $D^{+}_{m,n} = \sup[F^{*}_{m}(x) - G^{*}_{n}(x)],\ \forall x$. Here we are testing the alternative, $G(x) \leq F(x),\ \forall x$ and $G(x) < F(x)$, for some x with rejection region $D^{+}_{m,n} \geq D^{+}_{m,n,\alpha}$. Likewise, the "lesser than" KS test removes the absolute value function and changes the order of the EDFs. This is defined as: $D^{-}_{m,n} = \sup[G^{*}_{n}(x) - F^{*}_{m}(x)],\ \forall x$. Where we are testing the alternative, $F(x) \leq G(x),\ \forall x$ and $F(x) < G(x)$ for some x, with rejection region $D^{-}_{m,n} \geq D^{-}_{m,n,\alpha}$ [21].

When we are comparing the values for the coefficient of variance, we will use the "lesser than" test since our hypothesis is such that we expect the columns where fixations occurred to have higher CV values; thus we expect the EDF weighted with the n_j values to lie beneath the EDF from a uniform distribution. We will use the "greater than" test for the distances to doors and windows since we expect those columns that have lesser distances to have a greater number of fixations; we expect the EDF that is weighted by the number of fixations to lie above the EDF that is uniformly distributed.

Our two distributions will be the vector of repeated values according to the number of fixations (F) and the distribution of the values for all columns (G). The F^{*} are the EDFs shown in Figure 10. We must create a vector for the uniform distribution as done above, but we will use all of the columns and only take one sample from each of the independent variables. From this sample, we will sort the vector and create the EDF for the uniform distribution, or G^{*}_{n} values. Figure 12 shows the two EDFs where the red points are the fixation EDF and the blue points are the uniform EDF.

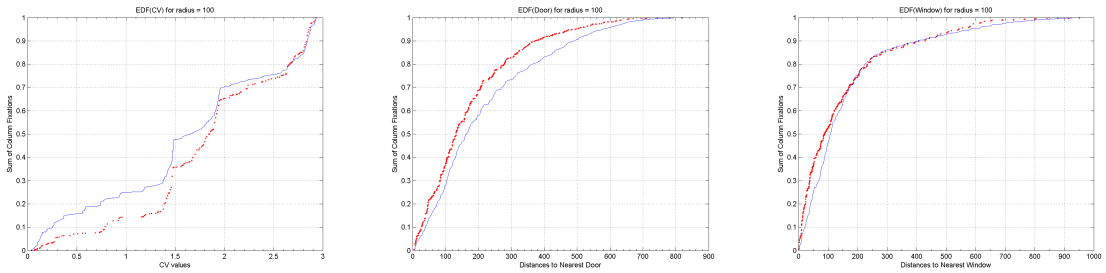

Figure 12: **Left plot:** EDFs for CV values. **Center plot:** EDFs for door values. **Right plot:** EDFs for window values.

The two EDFs for the CV values shown in Figure 12 show that more fixations have occurred along the columns where the CV is greater as identified in our alternative hypothesis. The important aspect of the CV EDFs is that there is a significant difference between the lower CV values (left side) for the two EDFs, but the two EDFs are closer as the CV values increase. The EDFs for the distance to doors and the EDFs for the distance to windows shows that more fixations have occurred along the columns with smaller distances. The EDFs for window distance show stronger results for distances of approximately zero to 175 pixels before being equal or even dipping below the uniform line, before improving again around 450 pixels. The results for the distances to doors along a column appear to have a stronger result for those values that are closer to zero than the distances to windows along a column.

Table 12 shows the p-values from conducting the one-sided KS tests for each of the independent variables. These values show that we reject the null hypothesis for any reasonable level of significance. For CV values this supports the idea that greater changes in depth draw attention and cause fixations. It also suggests that more fixations occur near doors and windows.

Table 12: KS test results for three independent variables

Independent Variable	type of test	p-value
cv_j	lesser than	1.0351e-006
d_j	greater than	9.9188e-005
w_j	greater than	1.1451e-006

D. POINTS OF INTEREST

Since the EDFs of the distances to doors and distances to windows are similar, what if we treat these two the same? A way to capture this is to simply treat doors and windows as "points of interest" (POIs) or "focal points." This is supported by previous work that shows that visual attention is drawn to "clutter" [11]. The idea of POIs can be traced to the what is called the "neo-classical" approach to search. Vaughn [3] outlines this idea and

shows that it can be implemented in models using a Markov process. This study will not address applying Markov processes to this method, but will leave that to further research. We will introduce this additional independent variable as dw_j; defined as the minimum value of the d_j and w_j. Table 13 shows this new definition that is added to Table 8.

Table 13: Definition of the POI variable for column vectors

Independent Variable	**Definition**
dw_j	$\min\{d_j, w_j\},\ \forall j$

When we treat these two variables as POIs, we achieve better results than when only dealing with the variables independently. Figure 13 shows the distribution, histogram, and the EDF plot with both the uniform EDF and the fixation EDF shown.

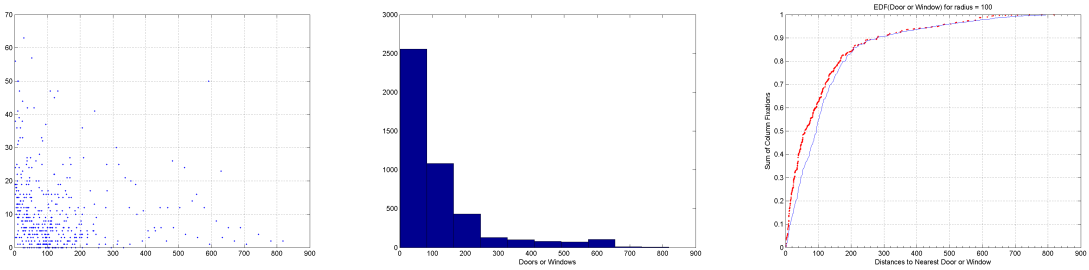

Figure 13: **Left plot:** Distribution of POIs. **Center plot:** Histogram of POIs. **Right plot:** EDF of POIs.

The p-value when combining the distances to doors and windows is 9.7e-009, which is lower than the p-values when calculated individually. The EDFs of the POIs show stronger results for distances to doors or windows along a column for pixel distances that are roughly between zero and 250. We do see better results for the rest of the values when compared to the EDF for windows alone; the fixation EDF dips below the uniform EDF line at only a few points compared to the plot shown in Figure 12.

The important aspect we are looking for here is nearness to a POI. As described in Chapter IV, the horizontal viewing angle for the participants was approximately $71.1°$. With an image with 1114 pixels across the horizontal axis, this equates to 15.67 pixels/de-

gree or 0.0638° per pixel. For 250 pixels this is equal to a 15.95° viewing angle between these POIs where fixations appear to be the same as a uniform distribution of the column information. Jungkunz [14] investigated the level of eccentricity, or the angular measurement between the center of a scene and a target, and the ability of an observer to detect targets. He used $0^{\circ}, 5^{\circ}, 9^{\circ}, 13^{\circ}$, and 17° as his classifications of eccentricity levels across the horizontal axis in his single target study, where eccentricities between 13° and 17° were classified as his greatest level of eccentricity. His results showed that there was significant increases in performance between targets located at a greater level of eccentricity and the number of fixations and the amount of time spent before detection. Using the KS test, we have the strongest results for the fixations that are less than this highest level of eccentricity and thus infer that a "point of interest" along a column does draw visual attention.

This third test has shown more support for confirming the first objective, even when aggregating the information along columns in a scene; i.e., search is still guided to salient scene information. We can also observe that combining the distances to windows and doors by combining them into POIs, improves the results of the KS test. Next, we must devise a method to implement these findings into Combat XXI and ACQUIRE in order to attack the second objective of this research.

THIS PAGE INTENTIONALLY LEFT BLANK

VI. APPLYING DISTRIBUTIONS TO COMBAT XXI

A. COMBAT XXI

The second objective of this study is to provide a method for Combat XXI to simulate how human unaided vision works in urban environments. Using the data from the previous study, we have shown above that search patterns of unaided search were strongly directed by salient scene information. We achieved similar results as the previous study when we removed a 100 pixel radius around targets using the Mann-Whitney U statistic, reinforcing the idea that eye is drawn to salient scene information. We also showed that aggregating the information along columns after removing a 100 pixel radius also gave us strong evidence that attention is drawn to the same scene information. With this in mind, how do we then translate the idea that salient scene information should drive search patterns into something that is more suitable to Combat XXI?

The general idea of this proposed model is to create a probability mapping that prioritizes which FOV the entity should search first by labeling each FOV with different levels of importance. A way to apply this method in Combat XXI is to assign varying probabilities, or percentages, of time by which to interrogate each FOV. A graphical example is shown in Figure 14, which shows a generic overhead view of a computer entity with equally divided FOVs within a larger FOR. Each FOV, FOV_i, is assigned a probability, p_i, to determine the amount of time to interrogate the FOV. The probability can also be thought of as a percentage of the total amount of time given for the entity to interrogate the FOV. A queue can be established based on decreasing values of p_i; the entity searches each FOV in the FOR, beginning with the FOV that has the greatest p_i value.

B. COMPARING POINTS OF INTEREST AND FIXATIONS

In order to satisfy the second objective, a saliency map was created for the POIs in each scene and compared to another saliency map based on the fixation data. In order

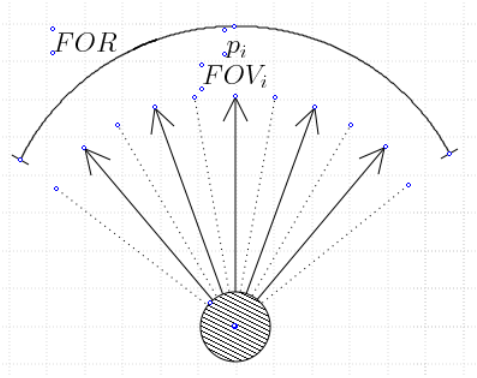

Figure 14: Overhead view of a computer entity

to do this, we used a simple Gaussian function of two variables. The use of a Gaussian function in a similar manner has been used by Dorr [4] and Torralba [10] although with different parameters. Equation VI.1 shows the two variable Gaussian function; where σ is the standard deviation, x_0 and y_0 are the fixation or POI locations, and x and y are mesh points.

$$f(x,y) = e^{-\left(\frac{(x-x_0)^2}{2\sigma^2}+\frac{(y-y_0)^2}{2\sigma^2}\right)} \tag{VI.1}$$

First we identified the POIs for each scene. This was conducted by two different viewers. The first was the author, who relying upon actual combat experience located the areas in the scene where he would search. The other was a Combat XXI expert, who placed POIs in locations that would align with possible POI locations within the actual simulation. Figure 15 shows an example scene with the points of interest indicated by blue circles.

Next, for each scene we created a Gaussian distribution centered at each POI. Let A_k be the 598×1114 matrices containing the values of the Gaussian calculated using equation VI.1 above for each POI in the scene, where $k = 1, ..., n_s$, where n_s is total number of POIs in scene $s = 1, ..., 16$. Each entry of these matrices is denoted by $f_k(x_j, y_i)$, where $i = 1, ..., 598$ is the corresponding row and $j = 1, ..., 1114$ is the corresponding column.

$$(A_k)_{i,j} = f(x_j, y_i) \tag{VI.2}$$

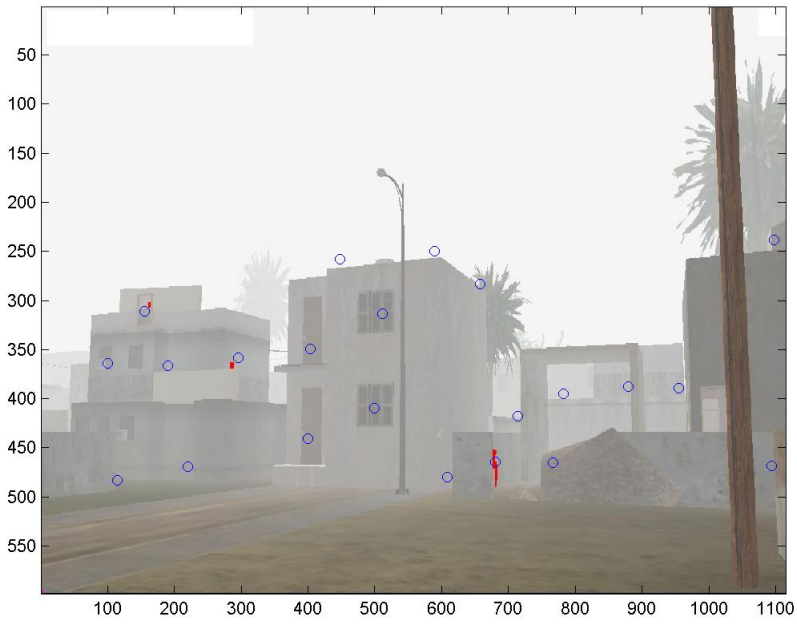

Figure 15: Sample scene with POIs

Each of the A_k for a POI are then summed for each scene. We will define the resulting matrix as B_s. In order to get a probability distribution, we must normalize the matrix B_s by dividing by the sum of all entries in B_s. We will define this matrix as C_s. The symbolic representations of these two steps are shown in equations VI.3 and VI.4.

$$B_s = \sum_{\forall k} A_k \tag{VI.3}$$

$$C_s = \frac{B_s}{\sum_{i,j}(B_s)_{i,j}} \tag{VI.4}$$

Since all of the entries in the matrix C_s sum to one, it can be used to represent the recommended probability distribution based on the locations of the POIs. This same process is then repeated, but using the fixation locations instead of the POI locations. We will define the resulting matrix as D_s to distinguish it from the POI matrix. An example scene's heat maps for POIs and fixations are shown in Figure 16

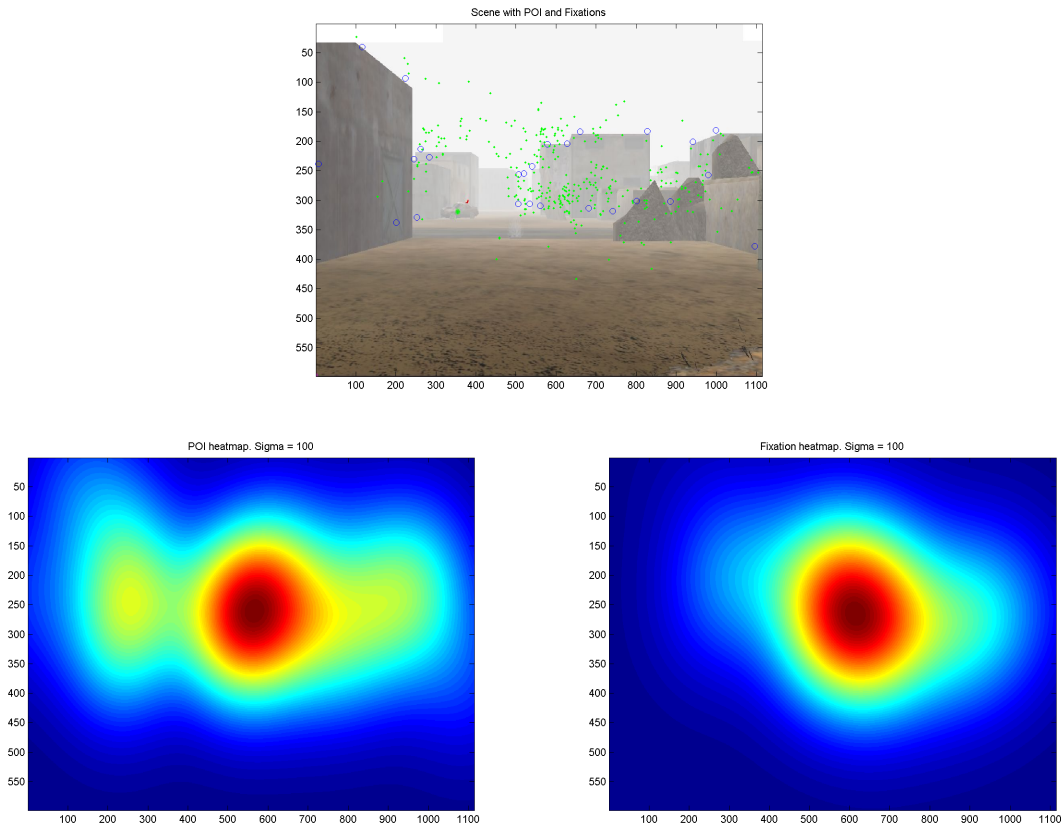

Figure 16: **Top plot:** Example scene with POIs and fixations. **Bottom left:** Heat map generated from POIs. **Bottom right:** Heat map generated from fixation points.

C. SELECTING SIGMA VALUES

By selecting different values for sigma, the probability distributions can vary significantly. If the value of sigma is too low, each POI is treated as a small area of interest. This will cause the heat map, and associated probability distribution, to consist of many small, sharp peaks. If the value of sigma is too high the heat map will generally have only one or possibly two peaks, where we can not easily differentiate different POIs. Figure 17 shows two heat maps from one scene where sigma is equal to 25 on the left and equal to 300 on the right.

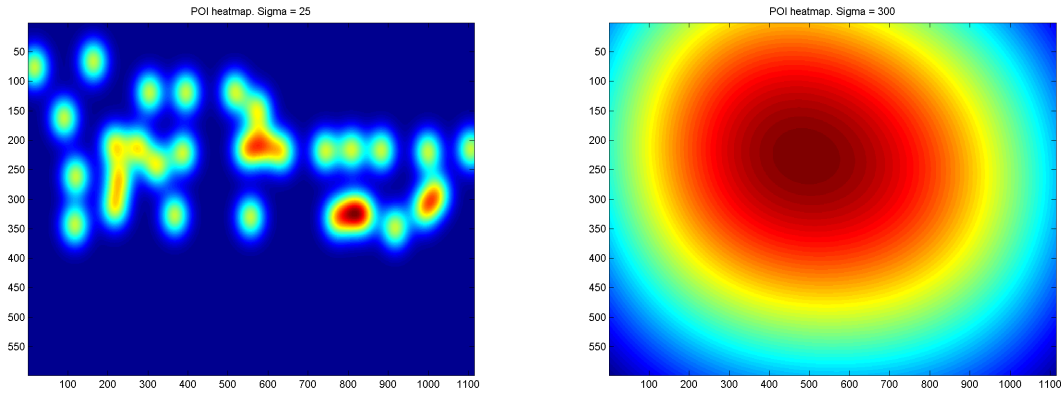

Figure 17: Example scene heat maps for two values of sigma. **Left plot:** Sigma = 25. **Right plot:** Sigma = 100.

After comparing various values of sigma, this study settled on a sigma value of 100 as a suitable estimate. This allowed most scenes to have two or more significant peaks in their heat maps. However, some scenes still had one large peak, such as shown on the right side of Figure 17. An example of using a sigma value of 100 for the same scene is shown in Figure 18 and also for a different scene in Figure 16.

If this method is adopted in Combat XXI in the future, additional research is required to optimize values of sigma that will better portray human vision. Possible ideas for research are to vary sigma according to the classification of the POI, such as whether it is a door, window, or edge along a building or rooftop. Distances to the POI could also dictate a suitable sigma value. A possibility for this is to assign lower values for extremely close

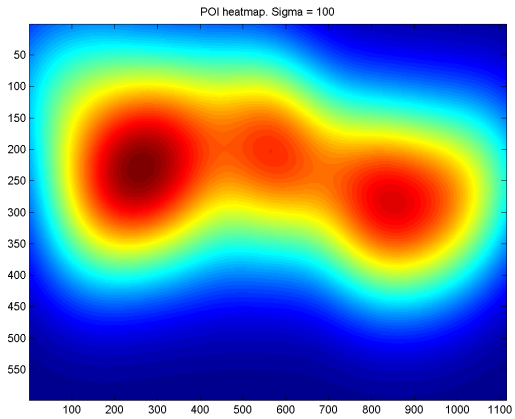

Figure 18: Example scene heat map for sigma equal to 100

or farther POIs and assign larger values for POIs with middle distances. The obliqueness, or angle of the face of an object, can also dictate varying weights.

D. USING COLUMNS / BINNING COLUMNS

As mentioned previously, ACQUIRE does not populate the FOV target queue with the vertical axis information. To account for this and to provide information that ACQUIRE can use, we will sum the total probability of C_s and D_s according to n bins along the horizontal axis. This will correlate to assigning ranks or values to multiple FOVs within a larger FOR. Figure 19 shows a generic scene and the evenly distributed columns according to the number of bins.

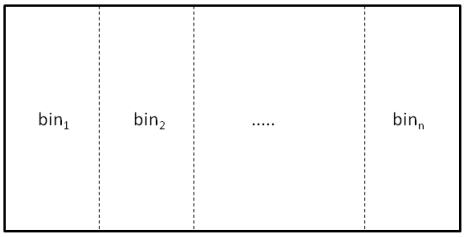

Figure 19: Generic scene with n bins along horizontal axis

Each bin gives the probability for looking in that particular direction where each direction has a FOV equal to $71.1°/n$. To determine how well the POIs predict the probability

of interrogating a specific direction, we must compare them to the probabilities determined with the empirical data. One way to do this is to compare the orders that are dictated by the probabilities. We let the highest probability among each bin be assigned to the queue in the first position. The next highest probability is assigned the second position. This is continued until the lowest probability is reached and it is assigned the n^{th} position in the queue.

E. DISCUSSION OF TWO METHODS

In order to determine if the POI distributions, C_i, can be used to dictate the direction of a computer entity's gaze, we must compare it to the empirical data distributions, D_i. Two methods were used in order to compare the POI distributions and the empirical data distributions. The first method uses the fixation data for all participants as a whole and compares it to the POI distributions for each scene. The second method uses the fixation data for each participant and compares it with the POI distributions for each scene. Both methods continue to use the removal of a 100 pixel radius around targets to remove the influence of any target's presence.

1. Method 1: Comparison by Scene

This method compares gaze patterns as determined by the POI distributions and the fixation distributions for each scene. As a preliminary step, the primary directions as dictated by the two distributions are compared. This step was executed multiple times by varying the number from two up to six bins. Table 14 shows the number of scenes where the POI distributions dictated the same primary direction as the fixation distributions. The first number corresponds to the number of scenes when using the author's defined POIs; the second number corresponds to the number when using the Combat XXI expert's POIs.

When the number of bins is two, the decision is for the entity to choose to look left or right as dictated by the probability mapping. We would expect that the primary direction should be the same for the POI distributions as the empirical data's distributions. As Table

Table 14: Number of scenes having a matching primary direction (out of 16 scenes)

Bins	# scenes w/matching ranks author | expert	FOV angle
2	11 | 16	35.55
3	9 | 11	23.7
4	7 | 9	17.77
5	8 | 10	14.22
6	7 | 10	11.85

14 shows, when the number of bins is two, the number of times that the POI distribution had the same primary direction as the empirical distribution is 11 out of 16 scenes for the author, but a match for all of the scenes when using the expert's POIs. For three bins it drops to 9 out of 16 scenes for the author and 11 out of 16 for the expert. When the number of bins is greater than three it drops to 50 percent or less when using the author's POIs, but stays above when using the expert's POIs. The expert had significantly more POIs per scene when compared to the author's POIs. The author chose the POIs strictly on important aspects of the scene as seen by a dismounted Soldier. Whereas, the expert was able to identify more POIs by having a thorough understanding of how the geometry is defined in Combat XXI and how the actual objects are defined in the simulation.

After checking the performance of the POI model with only the primary direction, it is important to check the performance of the model when predicting the two most important directions. Table 15 shows the results of the model compared to the empirical data. The second column shows the number of scenes that the POI model matched exactly the ordering as the empirical data. The third column shows the number of scenes that matched the first and second direction, but not necessarily in the exact order. Again, the first number in each column corresponds to the author defined POIs, while the second number corresponds to the expert's POIs.

Table 15 shows that when there were three bins, eight or nine of the scenes had the same exact two primary directions. It also shows that 12 or 14 of the scenes had the same two primary directions, while stating that the third direction was of lesser importance.

Table 15: Number of scenes having two matching directions (out of 16 scenes)

Bins	# w/exact ranks author | expert	# w/two primary ranks author | expert	FOV angle
3	9 | 8	14 | 12	23.7
4	5 | 7	12 | 13	17.77
5	4 | 7	8 | 11	14.22
6	5 | 5	6 | 6	11.85

Essentially, one of the directions was of such little importance compared to the other two that the POI distribution and empirical distribution were practically the same. There were also good results with four bins in finding the two most important directions. Results from the POIs from the author and the expert were relatively similar when checking for the best two directions, either matching the directions exactly or with the two primary ranks.

These results for both the primary direction the two primary directions for a larger number of bins was not as encouraging as hoped for. When looking at the primary direction, when there are two or three bins, the model does relatively well, doing better than what a random search would do; i.e., 50%. However, when the number of bins is greater than four, the model does not predict better than a random search pattern. Future research could exploit the finding of an optimal FOV size. As the number of bins increases each FOV decreases; however, if the FOV is too small the model may not properly represent search with the naked human eye.

The model here is comparing the probability mapping based on the POIs with the probability mapping based on the fixations. The fixations are generalized over the 20 seconds of the experiment and a general direction is determined to be the primary direction of search. This method has a drawback in that it determines the first direction based on the time of the whole experiment. Ideally, we would use the order in which the fixations occurred to determine the search pattern, but removing those fixations that were influenced by target information sacrificed any timing information regarding fixations and saccades.

2. Method 2: Comparison by Participant

The second method compares gaze patterns as determined by the two distributions, but is separated by participant and by each scene. The creation of each fixation distribution, D_s, for this method only used fixations from one participant for each scene. As in the first method, these distributions were then compared to the POI distributions. We define a matching rank for participants if the two distributions correlate for more than 50% of the scenes. It is important to note that the number of scenes will vary since the removal of fixations influenced by target presence may cause some scenes to have no fixations. These scenes are excluded from the calculation of the percentage of scenes.

The first test, while using this method, examined the primary direction with varying numbers of bins, similar to the first method. Table 16 shows the results of the number of participants, out of a possible 25, that correlated with the POI distribution. The left side of the second column shows the number of participants with matching ranks as the fixation data according to the author's POIs; the right side uses the expert's POIs.

Table 16: Number of participants having a matching primary direction (out of 25 participants)

Bins	# participants w/matching ranks author | expert	FOV angle
2	20 | 24	35.55
3	15 | 11	23.7
4	5 | 7	17.77
5	3 | 6	14.22
6	1 | 5	11.85

When there are only two or three bins, the performance of the POI distribution for finding the primary direction matches the empirical data very well. For example, when using the author's POIs and only two or three bins, 80% and 60% of the participants had matching distributions, respectively. When using two bins, the expert's POIs matched 24 out of 25 participants, which is significant; nearly all of the participants fixated to the side

of the scene where the POI distributions dictated. The performance quickly drops off when there are more than three bins, although the expert's POI locations tend to predict better than the author's POI locations. A possible reason for this decrease in performance with more than three bins will be discussed at the end of this chapter.

As with the first method, the performance of the POI distribution to determine the two most important directions are examined. Figure 17 shows the results when finding the two directions. The second column shows the number of participants with exact ranks and the third column shows the number of participants with first two primary ranks. Again, the left side of these columns are the results using the author's POIs while the right side uses the expert's POIs.

Table 17: Number of participants having two matching directions (out of 25 participants)

Bins	# participants w/exact ranks author | expert	# participants w/two primary ranks author | expert	FOV angle
3	2 | 3	24 | 24	23.7
4	0 | 0	9 | 12	17.77
5	0 | 0	3 | 8	14.22
6	0 | 0	0 | 0	11.85

The results from Table 17 are astonishing; when looking for an exact match, the number of participants' fixations that match the POI distribution is two with three bins and then quickly goes to zero matches for a greater number of bins. When only looking for the two best directions, almost all of the participants match up when the FOR is broken into three FOVs. This essentially says that there is one of the three directions that nearly all of the participants did not place much importance in and the POI model would dictate this direction as not important as well. The results for both the author's and the expert's POIs are essentially identical for this step. Beyond four bins though, the model did not predict the participants' empirical distribution well.

Why does an increased number of bins show decreased performance? The answer here could lie in the fact this method has reduced the number of fixations in a scene by only focusing on a particular participant. As we increase the number of bins, the number of fixations per scene is reduced even further. The sample size of fixations is reduced to such an extent that we do not have enough data to make a sound judgment.

The second objective of this study, to create a guided search method to implement unaided human search in Combat XXI, has showed some signs of promise. When the number of bins selected to break up a FOR is small, search can be guided to the general region of a scene that contains the most salient information. By simply prioritizing the search mechanism in ACQUIRE to search the FOV with the most important scene characteristics, we might theoretically improve the fidelity of search. More research is required to determine the optimal size of a FOV angle; an optimal size FOV, so long as it models human vision, could help improve search performance in ACQUIRE.

VII. CONCLUSIONS AND AREAS FOR FUTURE RESEARCH

A. CONCLUSIONS

The first objective was to utilize fixation data from the tier II experiment to determine whether or not salient scene information indicates where a human subject fixates. The results showed strong indications that this is true and that scene information is a driver in the method by which a human observer searches a given scene. The first test used the complete data set for the tier II data, where the data set was sanitized to remove fixations with times outside of the 20 second window, as well as fixations that did not register with the equipment used. Using the Mann-Whitney U statistic, as done in the previous study [1], similar results were achieved using a 28 by 28 grid overlaid on each scene. The U statistic value was 0.7802 with the full set of data compared to 0.79 as discovered earlier.

In order to focus in on the salient scene information and not on possible fixations caused by the presence of targets, the second test removed fixations and the discrete cells within various pixel radii around targets. All of the various radii showed improvements in the U statistic, indicating that the scene information does drive search. Based on the experimental setup and the foveal point of the human eye, a pixel radius of 100 was settled upon. The resulting U statistic value when removing target information of 100 pixels was 0.7875, showing more evidence that search is indeed guided.

The third test conducted, aggregated the scene information along the columns of the scenes. This test assumed that removing a 100 pixel radius around targets sufficed in removing the fixations and scene information. The scene information was aggregated in columns in order to extract information that ACQUIRE can more readily use. Doing so, however, presented a problem with using the Mann-Whitney U statistic. The number of benign instances was reduced to such a small number that it did not lend itself to using the U statistic. Instead, it was necessary to use the Kolmogorov-Smirnov two-sample test, which compares the EDF of the number of fixations in the discrete cells along the columns

of a scene with the uniform EDF for the scene. The p-values for the coefficient of variance and distances to POIs (doors and windows combined) was practically zero. The results showed that the observers' gazes were not uniform but driven by changes in depth and the presence of POIs.

The results from the three tests conducted with the tier II fixation data showed a strong relationship between where fixations occurred and salient scene information, echoing the idea that "search is guided" [8]. The results from the column aggregated data lead into the second objective; that of developing a probability mapping to dictate how a computer entity should conduct search. This will allow for a nearly direct translation in to the methods that ACQUIRE currently uses. In theory, there is no need to completely redesign ACQUIRE from the ground up, but we should expect improved fidelity by simply adding an additional layer that dictates search patterns of computer entities.

The second objective was to develop a probability mapping to dictate how a computer entity should conduct search in an urban setting. The tool used here created probability mappings based on the POIs in a scene and the fixation data. The first method compared the collective fixations of all the participants to the POI distribution for each scene. When looking at the primary direction, breaking the FOR into two FOVs yielded the best results, having all of the scenes matching the empirical data using the expert's POIs, and having 11 out of 16 scenes matching the empirical data using the author's POIs. Breaking the FOR into three FOVs resulted in 11 out of 16 scenes matching when using the expert's POIs. Beyond three bins though, there was no improvement over using a random sequence to the FOVs for the primary direction of search. When looking for the two primary directions, breaking the FOR into three FOVs yielded 9 matching scenes for an exact match, but showed greater improvements when trying to identify the two best directions with 14 out of 16 scenes. Using four bins showed 5 matching scenes with an exact match, but 12 matches when only looking for the two directions. Results for a greater number of bins were not as promising, often showing lesser performance than a random search.

The second method focused on the participants individually and measured the performance for each scene, comparing the the POI distribution with each of the fixation distributions for each scene. This method compared the primary direction and the two most important directions as well. When comparing the primary direction as dictated by the two distributions, using two or three bins showed great results. This can be used to show not where an observer searches, but where they do not search, which can be equally effective in describing search behavior. Breaking the FOR into two FOVs had 80% of the participants have matching ranks with the author's POIs, but nearly all of the participants with matching ranks. Beyond 3 bins though, the results quickly dropped off for the number of participants with matching primary directions. When comparing the two most important directions, performance in finding the exact two directions was very low. However, when only identifying the two most important directions, without an exact match, the performance for using three bins was much higher. Using three FOVs showed that nearly all of the participants, 24 out of 25, directed their gaze at the same general area.

The second objective was not met with as great success as the first objective. It does, however, show some signs of promise when using smaller numbers of bins. When breaking the FOR into only a few FOVs, the results were significantly better than when when using smaller FOVs. The results when trying to find the two significant directions, or FOVs, in a scene, indicate that we might want to change the question, "where should we direct search?" to "what areas of a scene are less salient, and thus can ignore or place lesser importance on them?" It is also speculated that the reduction in the sample size of fixations hindered this study. The study would have benefited from a larger eye tracking data set. The Mobile Eye tracking device allowed only for the tracking of the coordinates of the eye relative to the head, but did not track the eye coordinates relative to the scenes. The method used in the previous study worked well in identifying the fixations, but it did so at the cost of reducing the amount of information in the data set. The makers of the Mobile Eye device currently have technology that would allow the device to track the movements of the eye

relative to the scene parameters. Re-conducting a similar test with the newer technology would allow for a fixation data set that contained significantly more information.

The presence of targets in the majority of the scenes also hindered this study. The removal of targets allowed us to show evidence that search is guided, however, it did not allow for us to extract information about the lengths of saccades or the ordering of fixations. In order to get this information, the study would have to be conducted again with many of the scenes having zero targets. A sufficient study could still have targets present, but the number of scenes with targets would have to be significantly reduced.

The search model presented in this study can not take into account the ordering of fixations as mentioned above, but merely uses the fixations over the entire 20 seconds for comparison to the POI distribution. This problem was attempted to be circumvented by focusing on the first five or ten seconds of fixations, but this only caused the data set to be further reduced. This study also assumed the same sigma values when creating the saliency maps for the POIs as well as the fixations. Assigning different values for different categories of objects could improve the performance. Categorical search has been shown to perform better than a random search [7]. For example, sharp edges might be determined to be of greater importance than windows or doors. This could be true except when the obliqueness of a window or door is high. This would occur as an Soldier moves through an environment and is presented with windows or doors that they previously unaware of due to a change in position. The Soldier will need to interrogate this new object since it would be perceived as a possible threat.

B. AREAS FOR FUTURE RESEARCH

1. Preattentive Stage in Simulations

Much of the psychological literature regarding cognition and search agrees on the idea of a preattentive stage when conducting search. This stage extracts the necessary information from a scene and identifies what areas in the scene must be interrogated attentively. The simple windshield-wiper search pattern does not take this into account and is only

conducting a random search. The use of POIs to guide search appears to show some potential and could improve the performance of ACQUIRE. Precomputing saliency maps for locations in an environment would help reduce the overhead when running Combat XXI and could represent the preattentive stage of human vision. A boost in performance could be realized if we simply order the FOV queue in ACQUIRE according to their importance, even when using the windshield wiper scan pattern within each FOV.

The preattentive stage could also allow for the presence of targets by using the features of a target when conducting search. If a target's features are considered to be easily noticeable, the location of the target can be assigned a greater value when creating the saliency map. If a target is not easily noticeable, search would be conducted as normally dictated by the saliency map. The idea of a relevance map could also be combined with the saliency map to improve the representation of human search [10, 14]. The relevance map is created similarly to the saliency map, but is generated by identifying possible target hiding positions. For example, if we are looking for a vehicle, we ignore many areas in a scene where it is impossible for the vehicle to be located. If we are looking for a human target we may have more locations to interrogate than when searching for a vehicle threat, but would not have to interrogate open sky or featureless sides of buildings unless there is an easily identifiable target.

Another aspect of human search, is that often observers will visit the same location, but often not immediately after inspecting the location. They will usually inspect other areas in a scene before returning to an area that was "previously fixated and found not to be targets" [4]. With the current FOV target candidate queue in ACQUIRE, a previously searched FOV does not get revisited. If no target was found in the initial search, it is assumed that a target will not be present in a future search. Adding multiple searches of the same FOV in the search algorithm could allow for better search performance as well as possible detection of moving targets in ACQUIRE. This study did not examine the revisiting of a FOV; future studies could examine this aspect and its application to improving the current model.

2. Soldier Training

Understanding the human search process can also improve the training of Soldiers. Doll and Home observe, "after extensive practice, military observers are often able to immediately pick out targets in cluttered scenes that novice observers must search for painstakingly" [11]. They are speaking of an idea called "pop-out," or the ability to preattentively process a scene quicker when looking for a particular type of target or threat. An example of this type of training would be putting Soldiers in a realistic simulation where they must identify a threat and then act accordingly. The Army currently uses a system called the Engagement Skills Trainer (EST) 2000 which has a shoot/don't shoot training module [26, 27]. The Soldiers watch a series of videos where they may be presented with a threat such as a man firing a weapon on them. A Solider participating in this type of training gains experience after repeated uses, where his ability to have the identification of a threat "pop-out" quicker.

The problem with the EST 2000 is that there are only a limited number of systems and they are costly to acquire and operate. One proposal is to use a smaller system that does not require a large amount of resources, but still can train the Soldier to increase his ability to have threats pop-out quickly. Alt **(need citation)** has worked on a similar system that trains Soldiers to identify Improvised Explosive Devices (IEDs) using real imagery. They do this by using mouse clicks to specify the locations where they think a possible IED may be emplaced. They are given immediate feedback to the actual location of the IED if there is one in that particular scene. In working with the system, they gain experience searching for IEDs without having to be exposed to the danger of an IED. The idea is that they will be able to preattentively process a scene quicker and have a quicker reflex of pop-out.

With some modifications, this same type of system can be applied to searching a scene for possible threats that a dismounted Soldier might experience. The Soldier would gain experience in the situations where a positive identification of a threat is necessary. By increasing the ability of a Soldier to preattentively process a scene by having a quicker ability for threats to pop-out, they can focus on other aspects of their mission. One main

drawback to a system like this, is the ability to interact with the environment with a real weapon system as they do with the EST 2000. However, it would still train the Soldier's ability to detect threats quicker, without having to experience the danger of an actual engagement.

Another aspect of training has to do with using the gaze patterns of experts in order to increase performance of novices. Dorr stated, "recording the gaze patterns of experts and applying it to novices, we can evoke a sub-conscious learning effect" [4]. His study examined the use of drawing attention to certain areas in a scene by using very quick flashes or blurring areas of a scene based on the patterns of an expert. The idea is that you will train a novice observer to examine certain salient aspects of a scene in a particular order, or at the least, draw attention to areas of a scene where an expert has fixated. This essentially allows for a novice user to gain experience by placing them in a simulated environment so that they will perform better when presented with a real world situation. This would best be suited for inexperienced Soldiers at the onset of their military training, such as in basic training, or for all Soldiers as they prepare to deploy to a hostile theater of operations.

3. Final Thoughts

The proper modeling of human vision and cognition is imperative. The models are used to acquire future systems that will help our fighting forces maintain a tactical advantage and fight and win our nation's wars. The old adage, "garbage in—garbage out," applies to models of human behavior, especially when dealing with current defense acquisitions that are aimed at improving situational awareness on the battlefield. If a model of human cognition is done poorly and does not accurately portray human performance, much time and many dollars will be spent with no resulting product.

We must continue to improve the models we have or create new models that can accurately portray human vision, as it is the means by which we interact with our environment. Alt and Darken [28] used post-combat questionnaires from 27 Soldiers where all cited vision as their primary sense of identifying a threat during daytime engagements.

During night time engagements, the two primary senses were vision and hearing. Future models could seek to apply the effects of hearing into models as well.

LIST OF REFERENCES

[1] P. F. Evangelista, C. J. Darken, and P. Jungkunkz, "Modeling and integration of situational awareness and soldier target search," *Journal of Defense Modeling and Simulation*, (in press).

[2] P. Evangelista, I. Balogh, C. J. Darken, and J. Ruck, "Visual awareness in combat models," in *The 20th Behavior Representation in Modeling & Simulation (BRIMS) Conference*, 2011.

[3] B. Vaughan, "Soldier-in-the-Loop Target Acquisition Performance Prediction Through 2001: Integration of Perceptual and Cognitive Models," tech. rep., Army Research Lab Aberdeen Proving Ground, MD, Human Research and Engineering Directorate, 2006.

[4] M. Dorr, M. Böhme, T. Martinetz, and E. Barth, "Predicting, analysing, and guiding eye movements," in *Neural Information Processing Systems Conference (NIPS 2005), Workshop on Machine Learning for Implicit Feedback and User Modeling*, 2005.

[5] L. Itti, C. Gold, and C. Koch, "Visual attention and target detection in cluttered natural scenes," *Optical Engineering*, vol. 40, no. 9, pp. 1784–1793, 2001.

[6] A. Treisman, "Preattentive processing in vision," *Computer vision, graphics, and image processing*, vol. 31, no. 2, pp. 156–177, 1985.

[7] H. Yang and G. Zelinsky, "Visual search is guided to categorically-defined targets," *Vision research*, vol. 49, no. 16, pp. 2095–2103, 2009.

[8] G. Zelinsky, "A theory of eye movements during target acquisition.," *Psychological review*, vol. 115, no. 4, p. 787, 2008.

[9] A. Treisman, "Features and objects in visual processing," *Scientific American*, vol. 255, no. 5, pp. 114–125, 1986.

[10] A. Torralba, A. Oliva, M. Castelhano, and J. Henderson, "Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search.," *Psychological review*, vol. 113, no. 4, p. 766, 2006.

[11] T. Doll, "Lessons Learned in Developing and Validating Models of Visual Search and Target Acquisition," tech. rep., Georgia Tech Research Inst. Atlanta Electro-Optics Environment and Materials Lab, 2000.

[12] L. Zhaoping and P. Dayan, "Pre-attentive visual selection," *Neural Networks*, vol. 19, no. 9, pp. 1437–1439, 2006.

[13] R. Klein, "On the control of visual orienting," *Cognitive neuroscience of attention*, pp. 29–44, 2004.

[14] P. Jungkunz, *Modeling Human Visual Perception for Target Detection in Military Simulations*. PhD thesis, Naval Postgraduate School, Monterey, CA, 2009.

[15] L. Harrington, "Adding urban search to traditional search within combat simulations," Tech. Rep. TR-2009-XX, U.S. Army Material Systems Analysis Activity, January 2009.

[16] E. Jones and C. Lai, "Field of regard search in urban operations," Tech. Rep. TR-2007-37, U.S. Army Material Systems Analysis Activity, November 2007.

[17] E. Grove, "Validation of the search and target acquisition (sta) time-limited search model for target detection," Tech. Rep. TR-731, U.S. Army Material Systems Analysis Activity, December 2003.

[18] Applied Science Laboratories, *Operation Manual: Mobile Eye*, version 1.35 ed., June 2008. valid for EyeVision Software versions up to v.2.2.6.

[19] P. Evangelista, *The unbalanced classification problem: Detecting breaches in security*. PhD thesis, Rensselaer Polytechnic Institute, 2006.

[20] E. L. Lehmann, *Nonparametrics: Statistical Methods Based on Ranks*. San Francisco, CA: Holden-Day, Inc., 1975.

[21] V. K. Rohatgi and A. E. Saleh, *An Introduction to Probability and Statistics*. Wiley, second ed., 2001.

[22] A. Van der Vaart, *Asymptotic statistics*. Cambridge Univ Pr, 2000.

[23] M. Stephens, "Use of the Kolmogorov-Smirnov, Cramér-Von Mises and related statistics without extensive tables," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 32, no. 1, pp. 115–122, 1970.

[24] MATLAB, *version 7.8.0.347 (R2009a)*. Natick, Massachusetts: The MathWorks Inc., 2009.

[25] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical recipes in C*. Cambridge Univ. Press Cambridge, 1992.

[26] Headquarters, Department of the Army, *FM 17-12-8,Light Cavalry Gunnery*, February 1999.

[27] Headquarters, Department of the Army, *TC 7-21.10, Infantry and Weapons Company Guide to Training Aids, Devices, Simulators, and Simulations*, July 2009.

[28] J. Alt and C. J. Darken, “A reference model of soldier attention and behavior,” tech. rep., Naval Postgraduate School, Monterey, CA, Department of Computer Science, 2008.

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

3. Dr. Carlos Borges
   Naval Postgraduate School
   Monterey, California

4. MAJ Paul Evangelista
   TRADOC Analysis Center - Monterey
   Monterey, California

5. COL Michael Phillips
   United States Military Academy
   West Point, New York

6. COL(R) Lawrence Shattuck
   Naval Postgraduate School
   Monterey, California

7. LTC Dave Hudak
   TRADOC Analysis Center - Monterey
   Monterey, California

8. CPT James Starling
   United States Military Academy
   West Point, New York